THE ARCLIGHT GUIDEBOOK TO **MEDIA HISTORY** AND THE DIGITAL HUMANITIES

Charles R. Acland and Eric Hoyt, Editors

The Arclight Guidebook to Media History and the Digital Humanities is published online in e-book format as a collection by **REFRAME Books** (http://reframe.sussex.ac.uk/reframebooks) in association with **Project Arclight** (http://projectarclight.org).

REFRAME is an open access academic digital platform for the on-line practice, publication and curation of internationally produced research and scholarship. It is supported by the School of Media, Film and Music, University of Sussex, UK.

e-ISBN 978-0-9931996-7-7 (PDF)
e-ISBN 978-0-9931996-6-0 (ePub)

Copyright chapters © 2016 Individual Authors.
All Rights Reserved.

Copyright e-format & layout © 2016 Project Arclight.
All Rights Reserved.

Cover illustration and book design: MAT3RIAL

For REFRAME Books:
Managing Editor: Catherine Grant (University of Sussex)
Contact: reframe.us@gmail.com

THE ARCLIGHT GUIDEBOOK TO MEDIA HISTORY AND THE DIGITAL HUMANITIES

Charles R. Acland and Eric Hoyt, Editors

REFRAME Books
http://reframe.sussex.ac.uk/reframebooks
In Association with **Project Arclight**
http://projectarclight.org

THE ARCLIGHT GUIDEBOOK TO MEDIA HISTORY AND THE DIGITAL HUMANITIES

TABLE OF CONTENTS

Acknowledgements viii

A Guide to the Arclight Guidebook
Eric Hoyt, Kit Hughes, and Charles R. Acland 1

PART I: SEARCHING AND MAPPING 30

1. The Quick Search and Slow Scholarship: Researching Film Formats
Haidee Wasson 31

2. Search and Re-search: Digital Print Archives and the History of Multi-sited Cinema
Gregory A. Waller 45

3. Using Digital Maps to Investigate Cinema History
Laura Horak 65

4. Field Sketches with Arclight: Mapping the Industrial Film Sector
Kit Hughes 103

PART II: APPROACHING THE DATABASE 131

5. Low-Tech Digital
Charles R. Acland 132

6. Excavating Film History with Metadata Analysis: Building and Searching the ECHO Early Cinema Credits Database
Derek Long 145

7. Show Me the History! Big Data Goes to the Movies
Deb Verhoeven 165

8. How Is a Digital Project Like a Film?
Miriam Posner 184

PART III: ANALYZING IMAGES, SOUNDS, WORDS 195

9. Coding and Visualizing the Beauty in Hating Michelle Phan: Exploratory Experiments with YouTube, Images, and Discussion Boards
Tony Tran 196

10. Looking for Bachelors in American Silent Film: Experiments with Digital Methods
Lisa Spiro 216

11. Terminological Traffic in the Movie Business
Charles R. Acland & Fenwick McKelvey 238

12. Digital Tools for Film Analysis: Small Data
Lea Jacobs & Kaitlin Fyfe 249

13. The Slices of Cinema: Digital Surrealism as Research Strategy
Kevin L. Ferguson 270

PART IV: PROCESS, PRODUCT, AND PUBLICS 300

14. Digital Tools for Television Historiography: Researching and Writing the History of US Daytime Soap Opera
Elana Levine 301

15. When Worlds Collide: Sharing Historical Advertising Research on Tumblr
Cynthia B. Meyers 318

16. Networking Moving Image History: Archives, Scholars, and the Media Ecology Project
Mark Williams 335

17. Curating, Coding, Writing: Expanded Forms of Scholarly Production
Eric Hoyt 347

Keywords and Online Resources
Robert Hunt and Tony Tran 374

Contributors 389

Index 394

ACKNOWLEDGEMENTS

This book would not exist without the contributions of many institutions and individuals. The editors would like to thank them here.

Our work on Project Arclight: Analytics for the Study for Twentieth-Century Media was enabled thanks to an international Digging into Data grant, coordinated by the NEH Office of Digital Humanities and sponsored by the United States' Institute of Museum and Library Services and Canada's Social Sciences and Humanities Research Council. Additional funding came from the University of Wisconsin–Madison's Office of the Vice Chancellor for Research and Graduate Education, University of Wisconsin-Madison's Communication Arts Hamel Family Grant, and Concordia University's Research Chair in Communication Studies, specifically through the Media History Research Centre (MHRC). Additionally, we wish to thank the Media History Digital Library and its founder and director, David Pierce, for participating in Project Arclight. We are grateful to all of these institutions for believing in our work.

Receiving these grants was transformative for our work, but we needed great people to help us put our plans into action. At the University of Wisconsin–Madison, three graduate students—Kit Hughes, Derek Long, and Tony Tran—were instrumental to every phase of the project, from conceiving of the Arclight software, to enhancing the digitized searchable collections, to publishing the results of the project (including within this book). Kevin Ponto, Alex Peer, and Peter Sengstock made important contributions to the technical development and implementation of the Arclight software. We also benefited from the input of a broader set of Madison collaborators, especially Lea Jacobs, Michele Hilmes, Kaitlin Fyfe, Lee Konrad, Peter Gorman, and Carrie Roy. At Concordia University, Fenwick McKelvey, Elena Razlogova, Louis Pelletier, and Charlotte Fillmore-Handlon all worked to shape, build, and

test Project Arclight. They all provided essential critical and technical skill to the development, design, and dissemination of the project. Other local contributing scholars include Sandra Gabriele, Jason Camlot, and Michael E. Sinatra, and we thank them. In May 2015, Concordia University also hosted Arclight Symposium, which featured presentations by several of the scholars featured in the book. Our thanks go to Rachel MacNeill and Tyler Morgenstern for expertly organizing this event, as well as for other research assistantship.

We shared a prototype of the Arclight software at the symposium, but we were also grateful to share our work and receive feedback at several other conferences and events. Thank you to the conference organizers and audiences at the Society for Cinema and Media Studies, HASTAC, IAMHIST, IEEE Big Data, Film & History, University of Pennsylvania, Stockholm University, Ghent University, Film Studies Association of Canada (Ottawa), and Digging into Data (Glasgow). Your comments and questions prodded us to make our work better. Notable in this respect were Robert Allen, David Berry, and Ryan Cordell. Special thanks goes to Gregory A. Waller, Haidee Wasson, and Paul S. Moore for their feedback and advice throughout the process.

We feel fortunate to be publishing this collection with REFRAME Books, edited by Catherine Grant, whose enthusiasm, work ethic, and commitment to open access publishing are truly inspiring. The comments received during the peer review process contributed greatly to the shaping of the volume. And special appreciation goes to Charles Tepperman, whose exacting read and recommendations improved the work immensely. We are also grateful to be working with the outstanding Montreal-based design firm MAT3RIAL, who created our logo, website, and book design. Our thanks also go to Robert Hunt, Tony Tran, and Elise Cotter for their hard work and diligence in copyediting and preparing the final manuscript. And thank you to Emily Hagenmaier for her

support throughout the process.

Finally, we would like to thank all the scholars who generously agreed to contribute chapters to this book. We have learned a great deal from reading and interacting with your chapters. We are glad to now give readers of *The Arclight Guidebook* the opportunity to do the same.

A GUIDE TO THE ARCLIGHT GUIDEBOOK

Eric Hoyt, Kit Hughes, and Charles R. Acland

From the tremendous video libraries of YouTube and the Internet Archive to the text collections of the HathiTrust and the Media History Digital Library, media historians today confront the challenge of engaging with an abundance of cultural works and archival materials. For those invested in the digital humanities (DH), this abundance presents an opportunity to transform these materials' availability into data to be studied using a variety of methods. The primary point of departure for *The Arclight Guidebook to Media History and the Digital Humanities* is the exploration of this developing scholarly context. What new skills, competencies, and tools do media historians and scholars need in an era of digital research? What forms of publication and dissemination can and should we work in? And, perhaps most profoundly of all, what questions about media and culture can and should we be asking? Are we innovating and adapting digital tools to address our research questions? Or are we adapting our research interests to fit the available datasets and tools?

This book seeks to answer these questions—and raise new ones—by examining what media historians are doing right now with digital tools and methods. Across seventeen chapters, our contributors discuss the ways in which they are using or building digital technologies, assessing strengths and weaknesses, and responding to successes and failures. Some of the contributors share innovative methods and projects (e.g., Kit Hughes on scaled entity search or Kevin L. Ferguson on digital transformation using slicing and stacking). Others explore their engagement with existing technologies and methods (e.g., Cynthia B. Meyers' use of Tumblr as a dissemination platform or Haidee Wasson's reflections on word processing software and digital point-and-shoot cameras as transformative research tools). Ultimately, these projects seek to better

understand how media operate within processes of meaning making and cultural circulation. While some take on the central commercial media industries of the twentieth century (Hollywood, broadcasting, advertising), others examine nontheatrical and internet video production (industrial film, YouTube). All attempt to be reflexive about how the media of the twenty-first century—archives' content management systems, databases, video editing software, photography, and digital maps—shape our engagement with the past. By aggregating these perspectives, this collection seeks to be a "guidebook" that surveys what media historians are doing with digital tools and charts a course for how we might best move forward, especially in areas not covered by this collection, including video games and popular music.

(INTER)DISCIPLINARY RESEARCH CONTEXTS

As our book's title makes clear, we want to bring the highly heterogeneous fields of media history and the digital humanities into greater interaction. By some metrics, there has been minimal engagement between these fields. The annual DH Conference, organized by the Alliance of Digital Humanities Organizations, has tended to have far fewer participants from film and media studies departments compared to our peers in literature, linguistics, and classics. The DH Conference brings together an international community of scholars, librarians, and software developers who are at the forefront of innovating digital tools and methodologies for the humanities. Much like being a recipient of a grant sponsored by the NEH Office of Digital Humanities, participation in the selective DH Conference can provide a stamp of legitimacy that distinguishes insiders from outsiders in the digital humanities world. For years, the vast majority of media studies scholars, including many interested in digital media and technology, were content to sit on the outside.

There are reasons for this divide. Literary scholars benefited from (and contributed to) the early digitization of key text collections,

and they were able to leverage open source technologies for text analysis that were faster and more sophisticated than forms of automated moving image and sound analysis. Moreover, as a field that has been continuously "in crisis" for decades, maintaining relevance has been an obsession for literature departments. Turns to the contemporary media and technological environment have become a particularly robust effort to bolster that relevance, one product of which has been a commitment to what we now call "digital humanities." Media studies venues, in contrast, have not felt that special need, having the contemporary media and technological scene as foundational to their curricula and research.

However, the initial attachment of "digital humanities" to literature departments is changing. Due to the interventions of a new wave of scholars—and the expansion of broadband, blogging software, smart phones, and other networked technologies—the realm of practice that constitutes "digital humanities" has expanded.[1] Moving beyond the humanities computing tradition, today's DH includes new media criticism, digital publi-shing, innovations in peer review and scholarly communication, and forms of digital art production. Setting the boundaries of "the digital humanities" has become something of a cottage industry, with definitions spilling out of and across anthologies, conference talks, blogs, Twitter, and other forums.[2] Rather than take digital humanities as a circumscribed field of research, pedagogy, and outreach, we understand DH as a strategically deployed term of mutual recognition that enables contemporary knowledge workers to signal a shared project interested in the relationship between digital technologies and humanities work. Of course, disagreement exists over what that project is. In a sense, we are all digital humanists. Article databases, online catalogues, search algorithms, word processing software, email, and course management systems already shape contemporary academic work in countless ways. However, proportionally few place the question of how digital tools help constitute our questions, projects, and the process of

research and dissemination toward the center of their work. For those who do, taking up the mantle of DH signals their interest (to other researchers, to funding bodies) in entering conversations about what these relationships might mean for contemporary humanities practices. This is how we use DH here: to join a lively debate on what to do with the relationship between digital tools and humanities questions.

While keeping our understanding of DH as broad as possible, we hope that pointing out several productive strains of DH practice—many of which overlap—will help make the constellations of practices laying claim to the term legible to uninitiated readers. One of the first and largest areas of DH research concerns the transformation of information and objects into datasets that can be analyzed using computers. Books, film credits, trade journals, geographical atlases, and many other texts have been entered into databases, made into machine-readable lists, or input into simulations. This, in turn, enables computers (via algorithms or simple quantitative processes like counting words) to identify patterns invisible to the relatively small scale of human capacity. Visualization tools including graphs, word clouds, and interactive networks help make data meaningful both to researchers and their intended audiences. Arclight, a text-mining tool described below, offers an example of this avenue of DH research.

A second prominent area of DH practice seeks to digitize analogue materials to make them legible not to computers, but to a wider audience of human users. Libraries, archives, museums, amateurs, and others have sought to take advantage of the distributive capabilities of the internet. Examples range from online archives like the Texas Archive of the Moving Image to funding competitions such as the Council on Library and Information Resources' program, *Digitizing Hidden Special Collections and Archives: Enabling New Scholarship through Increasing Access to Unique Materials*. Eric Hoyt discusses additional collec-

tions projects in his chapter.

Similar to these digitization projects is a third strain of DH work that uses digital technologies to push the boundaries of what it means to write, publish, and consume scholarship. Some of these projects share the aforementioned concern to expand readers' or users' access to the products, objects, and tools of humanistic inquiry; *The Arclight Guidebook*, as an open access anthology, is a case in point. Others use the multimodal capabilities of digital technologies—which might combine video, hyperlinking, geographic information systems (GIS) and global positioning systems (GPS), text, interactivity, sound, images, and attention to temporality and duration—to create new kinds of arguments and experiences. We might think of this as Hayden White's "historiophoty" intensified. Ghosts of the Horseshoe, designed by Heidi Rae Cooley and Duncan Buell at the University of South Carolina, offers an example. Described as a "mobile augmented reality application," Ghosts of the Horseshoe enables users to visualize the integral role that slave labor played in the development of the university. As users walk a historic portion of the University of South Carolina's campus, an iPad loaded with the program offers an interactive "window" to the past that overlays existing architectures with images and text. Mobile technologies thus allow for a site-specific, multimedia argument about the politics of space, visibility, and historical erasure that targets experiential and affective registers. While distinct from text-based e-publishing, these latter sorts of projects share an interest in using digital tools to expand readership beyond traditional institutional surrounds while building new relationships between and among the producers and consumers of scholarship.

Taking these relationships between scholars, students, and wider publics as its primary focus is a fourth mode of DH work. Indebted to the rise of social media, these DH applications seek to build communities through ongoing conversations and engage-

ment. Though blogging, Twitter, Tumblr, and other online services ostensibly offer a form of publishing, emphasis rests on dialogue and discussion of in-progress work rather than the formal or semiformal presentation of research. Public-facing projects like the "Day of DH"—an annual collaborative social media event that asks digital humanists to simultaneously document (via a shared website) "a day in the life of the digital humanities"—attempt to build the international DH community through joint projects while making the work and value of DH legible to broader audiences.[3] Projects like *Film Studies for Free*, edited by REFRAME Books' managing editor Catherine Grant, provide open access to a wide variety of high quality resources for diverse discipline-specific communities. Personal research blogs (Ted Underwood's *The Stone and the Shell*), hybrid online publishing forums (Digital Humanities Now), and interactive platforms like Twitter offer spaces of dialogue and debate for researchers interested in exploring the relationships between digital technologies and the humanities. Although it could be considered as a separate area of DH practice, digital pedagogy is invested in similar questions and supported by many of these same platforms and practices. *Hybrid Pedagogy*, for example, publishes articles and podcasts, sponsors in-person seminars and online courses, and describes itself as "a community, a conversation, a collaboration, a school, and a journal." Taking advantage of web 2.0 and interactive technologies, such projects hope to create new opportunities for ongoing, critical engagement with the humanities among dispersed students, publics, and other scholars.

The last facet of DH we'll mention in this briefest of primers examines how digital technologies affect everyday humanities practices like reading and research. N. Katherine Hayles' work on technogenesis, for example, suggests that digital environments characterized by intensifying quantities of information shape readers' orientation to reading.[4] Careful to avoid technological determinism, Hayles is most interested in how the materiality and

technological affordances of digital technologies shape and are themselves shaped by human practice. Reflexively examining how our own professional tools (computers, cameras, the internet) are taken up, altered, and abandoned likewise provides insight into the longstanding interests, traditions, epistemological assumptions, and methodological debates in the humanities.

Within this context, film and media historians are engaging with the digital humanities in a variety of ways. Pursuing the first avenue mentioned above, historians are using digital technologies to help organize and make accessible information that helps us better understand the past. In their *Arclight Guidebook* chapters, Charles R. Acland and Derek Long both discuss digital collections of film credits that they have helped build. Mark Williams chronicles how the Media Ecology Project has brought together communities of archivists, researchers, and software developers to expand access to cultural heritage and contribute metadata back to the archives in the process. And, as Laura Horak describes in her chapter "Using Digital Maps to Investigate Cinema History," film and media scholars are embedding available information into spatial contexts to find new ways of understanding film circulation.

Scholars are likewise developing computational approaches to studying the moving image. Yuri Tsivian's Cinemetrics and Jeremy Butler's Shotlogger represent two especially significant attempts to combine software development, statistical research, and networked communities to arrive at more precise understandings of the evolution of film and television styles. In *The Arclight Guidebook*, Tony Tran and Kevin L. Ferguson both share digital methods for transforming YouTube videos and digitized films into data visualizations, which hold the promise of uncovering patterns otherwise difficult to identify. And, in "Digital Tools for Film Analysis: Small Data," Lea Jacobs and Kaitlin Fyfe make the case for using digital editing software to carefully analyze films at the level of in-

dividual frames, cuts, and sync points. Although there is still much work to be done in developing video analytics tailored for film and media historiography, the work mentioned above makes it clear that scholars are using software to explore questions of style at levels ranging from macro to micro, quantitative to qualitative.

Empowered by the same nonlinear video editing software described by Jacobs and Fyfe, as well as by the online distribution platforms of YouTube and Vimeo, media critics and scholars are also producing works of "videographic criticism." One especially important development was the 2014 launch of *[in]Transition*, an open access journal that publishes outstanding video essays and offers resources to novices looking to experiment with the form. Most of the *[in]Transition* videos have focused on questions of film genre, authorship, and style (the first issue, for example, included videos about Italian neorealism, Orson Welles, and Ingmar Bergman). More recently, though, Kevin L. Ferguson's "Volumetric Cinema" video essay has shown how the moving image can be well suited for showing data-oriented experiments, and a special issue encompassing five videos responding to articles published in *Cinema Journal* highlights some of the rhetorical possibilities of the form.[5] Just as innovative as its embrace of the audiovisual form for making scholarly arguments, *[in]Transition* employs an open peer review system. This means that *[in]Transition*'s peer reviewers know the identity of an author when they evaluate a video essay, and they share their own identities when they recommend a work for publication either "as is" or with "minor amendments" (they can also reject submissions and suggest substantial revisions before resubmission). After the creator integrates the review comments into the video essay, *[in]Transition* publishes the video essay alongside the author's research statement and edited, signed versions of the peer reviews. The academic peer review process—frequently opaque and closed—is thus rendered far more transparent to the audience, who can contribute their own responses and participate in a dialogue by posting in the com-

ments section of any video essay. Thus the journal is designed not only as a means of presenting selected videographic work, but to create a context for understanding it—and validating it—as a new mode of scholarly writing for the discipline of cinema and media studies and related fields. *The Arclight Guidebook* only engages with video essays in passing. However, we encourage readers who want to learn more about this important and growing scholarly form to explore the videos, reflections, reviews, and resources that *[in]Transition* has collected and made openly available.

One of the strengths of the video essay is the form's directness—scholars cite and comment upon audiovisual productions by including excerpts from the films and television shows they are discussing. Pursuing a similar goal, some film and media scholars have built web-based projects that combine video clips, images, and hypertext. The digital form allows for these media-rich projects to present users with interactive features that are absent from video essays, books, and articles. Tara McPherson and Steve Anderson at the University of Southern California have been especially important in building this form of scholarly publication. Their online journal, *Vectors*, founded in 2003, paired scholars from a range of disciplines—including film and media studies—with designers and coders to create what Miriam Posner describes in her *Arclight Guidebook* chapter as "databased digital projects." More recently, McPherson and Anderson's team have developed the publishing platform Scalar to ease the building and circulation of such projects. All of these projects offer productive interventions into the third tradition of DH detailed above.

As the technical know-how required to build a digital project changes, with some tasks becoming easier and other more complex endeavors being pursued, the skills of communicating effectively in the digital form and engaging audiences of scholars, students, and publics are more vital than ever. It is possible that the same amount of time that goes into writing a book can go into

developing software or a digital project. But will audiences find it, spend time with it, cite it, care about it, or afford it the same legitimacy? How can creators of software and digital projects design in ways that maximize the impact and contribution of their work? And to what extent must we attend to changing the expectations and reading practices of scholars, students, public users, and administrators who see and evaluate new forms of scholarship?

By providing a forum for film and media historians to reflect on their ongoing work, *The Arclight Guidebook* joins a small yet growing body of literature that grapples with the research potential offered by digital methods. Three edited collections of essays exploring the intersection of media studies and digital humanities have preceded this one and offer interested readers a range of case studies and additional perspectives. The 2009 book anthology *Digital Tools in Media Studies*, edited by Michael Ross, Manfred Grauer, and Bernd Freisleben, brought together scholars from around the world to share their work in building and implementing software research tools.[6] That same year, *Cinema Journal* published an "In Focus" section on "Digital Scholarship and Pedagogy," edited by Tara McPherson, featuring six short essays reflecting on building digital projects and using digital technology in the classroom.[7] And, in 2012, the inaugural issue of *FRAMES*, edited by Catherine Grant, gathered thirty-nine scholars, students, and practitioners who addressed the question, "have film and moving image studies been 're-born' digital?"[8] Given the rapid pace of change in both digital technology and the academic institutional landscape, it is time, once again, to reflect on these issues. The open access editorial philosophy that guided *FRAMES*'s first issue very much informs our work on *The Arclight Guidebook*. We have chosen to distinguish our contribution by creating a space for conversation targeted specifically to those working within the fields of media history and historiography.

The Arclight Guidebook is itself the product of an international

years-long digital humanities media history initiative. Project Arclight: Analytics for the Study of Twentieth-Century Media won a Digging into Data grant funded by the Institute of Museum and Library Services in the United States and the Social Science and Humanities Research Council in Canada. This grant supported two years of software development for Arclight, an application that allows users to visualize how terms (e.g., directors, cities, stations) trend across the two-million-page corpus of the Media History Digital Library. Project Arclight additionally pursued developing a critical discussion about the intersection of DH and media history, providing online resources and essays on the topic and hosting related events, including a talk by Johanna Drucker on the history and future of DH at Concordia University. A major event was Arclight Symposium, a three-day conference held in Montreal in May 2015 that brought together film and media historians, digital humanities literary scholars, and big data critics. Roughly half of the essays in the *Guidebook* emerged from the symposium. In this spirit, the remainder of this chapter traces our own experiences as media historians—some of whom were neophytes to the digital humanities when we embarked on Project Arclight—as we strove to develop a useful digital humanities tool for media history. Much as we anticipate that this volume's chapters will provide readers with a sense of the rich opportunities available for critical digital humanities media history, we hope that the following narrative offers guidance to those interested in pursuing large-scale collaborative and infrastructure-intensive DH projects. In developing Arclight, we designed the software with certain assumptions about historiography, the needs of researchers, and what makes for a good digital humanities tool. As we discuss in the next section, the experience of developing the software immersed us in the worlds of big data and code. However, our process continues to be informed by the questions of film and media history that excite us the most.

PROJECT ARCLIGHT AND NEW DIRECTIONS IN MEDIA HISTORY

Compared to the relative novelty and amorphousness of DH, the study of media history may seem like a stable and coherent enterprise. Yet to assume so would be to miss the recent growth of two exciting subfields and an important change in historiographic focus. Over the last decade, the study of media industries and "useful" media have both grown exponentially. On the first count, scholars have increasingly sought to understand moving image media not only as an art form and cultural product, but as a key North American industry of the twentieth and twenty-first centuries.[9] Likewise, "useful" moving image studies has grown from a few scattered mentions in the scholarly literature to a booming subfield working to address texts, contexts, and practices that include sponsored films, classroom and workplace media, and military training methods. "The great unread" of media studies, "useful" media challenges the logics of canonization and emphasis within moving image history by pointing to long neglected but significant industries and practices.[10]

Taken together, the study of media industries and useful media represent an emerging understanding of the diversity of moving image culture and industry that has been rarely accommodated by traditional film history. Essentially, the contributions of cultural theory and cultural studies have truly begun to be taken up by the conventionally more text-and-art-oriented film studies, an impact that had already been regularized for media scholars. This change in historiographic focus—which places more emphasis on the institutions and audiences surrounding films and media programs than those objects themselves—extends Jon Lewis and Eric Smoodin's calls to "look beyond the screen" and Richard Maltby, Daniel Biltereyst, and Philippe Meers's conception of the "new cinema history."[11] Media industry studies, useful media, and new cinema history all ask what happens to our conception of "the movies" when we move outside the theater and the home,

beyond narrow notions of auteurism and Hollywood production, and past categories of film as art or entertainment. Not only do such questions require scholars to expand their conceptions of the film, television, and radio industries, they require new research methods, additional strategies for imagining industrial and cultural relationships, and a wider variety of sources and evidence—all challenges that digital tools and large online collections have the potential to meet. These were the conceptual seeds that first inspired Project Arclight.

We wrote the grant application for Project Arclight in 2013 with our research teams at Concordia University in Montreal and the University of Wisconsin–Madison. Our elevator pitch was that we would create Twitter analytics for media history. Much like analytic firms use Twitter to identify contemporary actors and TV shows that are "trending" in global popularity, we were interested in mining discussions of media content from a historic collection of film and media magazines. In pursuing this research agenda, we wanted to create a user-friendly web application that would make digital methods accessible to a wide range of users. While big data computing *can* require significant equipment and coding experience, we believed that the scale of the individual user's resources should not have to match the scale of their research questions. We wanted Arclight to enable the expert and nonexpert alike to run historical analytics and generate a variety of visualizations in the pursuit of their own questions.

When we got the news in January 2014 that we had received a Digging into Data grant, we were confronted with the simultaneously thrilling and intimidating challenge of putting our grand plans into action. If the priorities of cultural theory guided the first stage of Arclight's conception, then the debates surrounding tool building in the digital humanities informed the second stage. We wanted to build a tool that would appeal to a broad audience of nonprogrammers, who expect a fast and intuitive user experi-

ence. However, we also wanted to heed the calls of DH scholars who have insisted that tools should be transparent and open to interrogation and reflection.[12] Topic modeling software, utilized and discussed in Lisa Spiro's chapter, has been critiqued for being a black box even as it has been taken up as a major innovation by some DH scholars. Although Matthew Jockers and others have used topic modeling to generate large-scale analyses of literary patterns, some critics have objected to the fact that few humanists truly understand latent Dirichlet allocation (LDA), the complex probability theory that provides topic modeling's algorithmic backbone.[13]

In trying to balance the competing desires for user friendliness and transparency, we wound up turning to a technology that has been somewhat maligned within the digital humanities: search. Digital humanities literary scholar Stephen Ramsay has referred to keyword search as "that most primitive of procedures" of computational text analysis.[14] Matthew Jockers has called on digital humanities researchers to go "beyond search" and adopt less familiar digital processes.[15] One problem with search, as Jockers points out, is that it fails to direct us toward larger themes and patterns that go beyond our keywords. Ted Underwood has pointed out that full-text search can turn into a "Boolean fishing expedition" in which researchers run different groupings of keywords until, finally, they find results that validate their initial hypotheses.[16] And, in addition to all of these risks, we should point out that search can become just as much of a black box as topic modeling, especially in the way relevancy algorithms elevate certain objects in the results above others.

Yet, as Haidee Wasson reminds us in her chapter, searching "is fundamental to scholarship. As researchers we search. We search for evidence that confirms our thesis and hopefully for evidence that does not." Information retrieval researchers have highlighted the power of search and its ability to save time and serve a range

of information needs. Also, in contrast to the low adoption rate of most digital tools, search is widely used by scholars, students, and public users.[17] For this reason, any new insights that researchers develop in understanding the search process become doubly valuable; the critical lens that researchers develop for searching historical questions can transfer productively into other online search experiences. And, because search is used so widely across a variety of contexts, the open source software community has developed a fast, customizable, and well-documented search engine called Apache Solr. In developing Arclight, we wanted to leverage these many affordances of search to create a new digital tool scaled to big data.

At its most basic, Arclight allows users to track and compare word frequencies across a highly tailored collection of film and broadcasting trade materials. One can enter a single search term or ten thousand—there is no limit to the number of terms one can search with a single Arclight query. The application uses a modified version of keyword search (via Solr) that, instead of returning full-text results (all 2,357 of the actual pages that mention "steel" in *Business Screen*, for example), compiles search metadata (e.g., the *number of pages* in *Business Screen* that mention "steel" by year) and organizes that information via visualizations and a downloadable comma separated value (CSV) file. Although users can opt to obtain results based simply on the raw number of pages that contain a given term, another search option returns normalized values that indicate what percentage of total pages in a given year feature the user's search terms. For comparative analysis, normalized figures are essential.

Arclight's visualization and CSV features complement each other by appealing to two different needs during the search process. While both the visualization and the CSV (a spreadsheet-style file) are available within seconds after executing a search, the easy digestibility of visualizations allows users to respond quickly to re-

sults. Users can follow up surprises in initial inquiries with more exploratory or refined searches, incorporating Arclight's results as feedback into shifting hypotheses. The CSV file, while less immediately meaningful, accommodates larger entity lists and allows for more granular and precise analysis. The raw data it provides can also be used in developing additional visualizations, as Kit Hughes discusses in her essay. Together, the visualization window, the CSV file, and the ability to open up the underlying page hits in Lantern allow for a flexible and iterative process not unlike the comparative reading and rereading of sources that constitutes the long middle of any research project.

Insofar as Arclight describes word frequency counts over a large corpus, the project recalls Ngram Viewer, which allows users to track word frequency counts in several Google Books corpora over time. However, the heterogeneity, gaps in materials, and lack of transparency in the development of the Ngram Viewer corpus makes it impossible to use critically for media history work. The company built their initial corpora of about five million books in seven languages (about a third of Google Books' online holdings) based solely on the quality of their metadata and the reliability of their optical character recognition (OCR)—the process by which printed words on a page become machine-readable.[18] Beyond these parameters, the Ngram Viewer is a black box. While the vast scale of Google's project (now estimated at 6% of all published books) may work toward the possibility of making general, broad-based claims, without knowing precisely what books are included and which are not—and in what proportions—it is impossible to contextualize (or even delimit by subject) Ngram results adequately. This problem is compounded for subject researchers interested not in general linguistic processes or generic cultural formations, but in targeted areas of cultural production and activity.

Arclight addresses these difficulties by building on existing DH work—the MHDL—already tailored to the needs and interests

of media historians. Directed by David Pierce and Eric Hoyt, the MHDL eschews Google's gargantuan scale and generalist focus to offer a carefully curated but still wide-ranging collection of periodicals and print materials geared specifically toward publics (academics, hobbyists, students, artists) interested in media history. In varying proportions, the MHDL includes media industry trade papers, fan magazines, technical journals, government documents, and amateur magazines in the public domain. While anyone can access the full-page scans of these materials online, they also constitute the corpus for searches performed via Lantern and Arclight. Although this corpus constantly fluctuates due to ongoing efforts to build the collection, its two million pages already represent significant holdings in core trade publications (*Variety*, *Film Daily*, *Sponsor*), major nontheatrical papers (*Business Screen*, the *Educational Screen*), and long-running fan magazines (*Photoplay*, *Modern Screen*).

By scaffolding onto the MHDL, Arclight accomplishes several goals. First, this iterative approach to a large-scale DH project allows for the refining and reworking of ideas that normally occurs in the interstices between conference presentation, article, and book. Second, it allows us to conserve resources by making use of existing infrastructures. Third, and perhaps most important in this stage of DH in media history, it builds on many researchers' familiarity with the MHDL and its search engine, Lantern. Although, as described below, critical awareness of the MHDL's limitations remains vital to fully contextualizing Arclight's computational results, catering to users' existing competencies and comfort promises to open DH methods to a wider audience.

Besides the significant expense and time required to make texts available online, the MHDL's structuring limitation is copyright. The copyrights for many US pre-1964 trade, fan, and technical publications were not renewed, pushing these texts into the public domain and making them available for reuse. However, most

post-1964 US magazines and a much larger span of international magazines are protected by copyright and not yet part of the collection. Although allowing copyright to determine the shape of the corpus used by Arclight affects the results, it is vital for maintaining transparency, which can in turn help researchers properly contextualize and qualify their analyses. Again, Ngram Viewer provides an instructive counter-example. Although Google's service allows users to analyze word usage in books from 1500 to 2008, users cannot always access the underlying texts to see how words operate in context due to copyright restrictions. By including only materials that can be accessed in their entirety, the MHDL corpus ensures that users can be critically aware of exactly how their terms appear in context. Rhetorically, this system also emphasizes the importance of continual back-and-forth movement between close and distant scales of reading.

Digitization decisions made in the development of Arclight trade perfect accuracy for speed and scale, with the hope that the latter two properties help mitigate problems caused by the former. One of the major difficulties for text digitization projects is the accuracy of OCR, which is dependent on a number of factors, including the quality of page scans, font, and language.[19] While hand correction is possible, the labor time involved would so reduce the output of the MHDL that it would be unusable as a historical resource. The quality of OCR in the MHDL will continue to vary, especially as OCR technologies become more accurate. In response to this problem, Arclight results count only the number of pages that feature a searched term—rather than attempt to account for every single mention of a term—in the hopes that term redundancy on a single page can help smooth the impact of poor OCR. The MHDL's digitization protocols also favor speed in their decision to upload periodicals—in most cases—in files delimited by full years rather than by individual issues. Although this prevents more granular analysis, for example of month-level data, it allows the MHDL to scan and upload material at a faster (and less resource-

intensive) rate. Taking a cue from archivists—workers who have long learned to live with bulk as a constitutive challenge to their mission—Arclight aims for "more product, less process" in order to increase access and usability.[20]

Building software is a humbling experience. Something that you think will take a week can take six months, and, even then, be a disappointment. Yet there was another humbling moment in building Arclight: the ultimate realization that no software, however wonderful it might be, will ever solve conundrums of tool building in the digital humanities. It is not enough to develop technical processes and user interfaces to explore media history's data. What is equally important, if not more so, is to develop interpretive frameworks for analyzing the results. This is why understanding the strengths and limitations of the MHDL as a corpus, discussed above, is so important. Researchers need to think about the corpus they are analyzing in relation to the entities they are searching and the digital technologies, algorithms, and data structures that comprise the process. This relationship between the corpus, entities, and digital is key to the interpretive process of Scaled Entity Search (SES), which Kit Hughes demonstrates in her chapter, "Field Sketches with Arclight: Mapping the Industrial Film Sector."

The importance of interpretive frameworks and reflecting on how researchers actually use digital tools in their work prompted us to assemble *The Arclight Guidebook*. The participants who have worked on Project Arclight are, for the most part, active media historians. The research questions, reflections, and general understanding of what we need in an effective and illuminating digital research instrument emerge from this expertise. Moreover, we wanted to facilitate a conversation that was more about research experience and usefulness than our software. In doing so, we encountered research and perspectives from scholars who changed our thinking about media history and the digital humani-

ties, which not only helped us redirect elements of the Arclight app, its interface, and its streamline parameters for queries, but also pushed forward our goal of expanding a critical discussion about the new world of digital methods for media historians and scho-lars.

BOOK STRUCTURE AND CHAPTER DESCRIPTIONS

The chapters that follow are grouped into four sections. Although the theme of studying media history in the digital age extends through the book, each section gathers a group of scholars to reflect on their work and how it relates to a core set of questions.

The first section, "Searching and Mapping," explores what is gained and lost when media historians employ searching and mapping tools in their research. These chapters remind us that historiography always involves making (and, hopefully, challenging) assumptions about time and space. All of the chapters explore questions of film circulation, and three of the chapters focus especially closely on the North American nontheatrical film industry, representing something of a dossier on the topic. First, Haidee Wasson reflects on her ongoing research into portable projectors and how ordinary consumer-oriented digital tools, such as point-and-shoot cameras and the "Finder" function on her desktop, have become integral to her search process. Next, Gregory A. Waller shares his research into the multi-sited exhibition of films in the mid-1910s. Waller describes his workflow as one of "re-search"—an iterative process of searching a range of digitized newspaper, magazine, and book collections, with one result sometimes providing the seed for new keyword queries. In the third chapter, Laura Horak explores the opportunities and challenges of using geospatial software programs to investigate media history questions. To conclude this section, Kit Hughes synthesizes searching and mapping, utilizing the abovementioned Arclight app and SES method to map the twentieth-century industrial film sector.

The forms of searching and mapping that Waller, Horak, and Hughes all discuss are dependent upon databases, datasets, and indexes. The second section, "Approaching the Database," contains four chapters that self-reflexively examine what it means to build and use a database. Continuing the emphasis on nontheatrical film history, Charles R. Acland shares his work in developing the Canadian Educational, Sponsored, and Industrial Film Project (CESIF) and argues for the continued relevance of "low-tech digital" projects that expand our knowledge of media history. Whereas Acland and his collaborators manually entered the information stored in CESIF, Derek Long developed Early Cinema History Online (ECHO) by algorithmically restructuring a dataset, compiled decades earlier, of thirty-five thousand American films released from 1908 to 1920. As Long demonstrates in his chapter, ECHO can be used to retrieve credits information, but it can also be utilized for metadata analysis, exploring questions such as "what were the most prolific film companies during the 1910s?" In her chapter, "Show Me the History! Big Data Goes to the Movies," Deb Verhoeven reflects on the Kinomatics Project and what it means to think historically about an ongoing data stream of global movie showtimes—descriptions of events that have not yet occurred. And, in the final chapter of this section, Miriam Posner uses classical film theory to interrogate DH and finds that database-based digital projects, like films, ask contradictory things of the audience. For readers seeking a fuller understanding of databases at the technical level, both Posner and Long also offer clear, succinct descriptions about how different databases and data structures work.

The third group of chapters, "Analyzing Images, Sounds, Words," explores methods for analyzing media history using digital tools. The authors of these chapters introduce their methods and share some of the results of their work, including some failed experiments (indeed, acknowledging failures and learning from them is an important part of DH work). Tony Tran begins the section

by explaining how he combined image, video, and text analytics to study the videos of YouTube star Michelle Phan and the discussion forums populated by her fans and "Anti-Phans." Similar to Tran's mixed-methods approach, Lisa Spiro applies four modes of DH literary analysis—concordances, n-grams, topic models, and text analysis software—to the MHDL to explore how the figure of "the bachelor" was represented in American silent film history. Charles R. Acland and Fenwick McKelvey continue the exploration of text-based analysis by using the Arclight app to examine the ways in which industry terms (e.g., "box office," "contract," "hit," "flop") were comparatively employed by 1930s Hollywood trade papers and fan magazines.

The last two chapters of this section move from text-based analysis of film history into digital analysis of the films themselves. In "Digital Tools for Film Analysis: Small Data," Lea Jacobs and Kaitlin Fyfe share their work in using video editing software as an analytical tool. By closely studying the relationship between sound and image on a shot-by-shot and even frame-by-frame basis, Jacobs and Fyfe arrive at precise analyses of films and music videos and make the case that we should not overlook the power of digital tools for investigating "small data." Pursuing a very different method of moving image analysis than Jacobs and Fyfe, Kevin L. Ferguson presents his surrealist method of using ImageJ software to transform stacks of film slices into three-dimensional research objects. Ferguson's chapter, as well as others in this collection, makes it clear that the digital turn for media history need not represent a strictly quantitative one.

The final group of chapters, "Process, Product, and Publics," attends to the workflows of researching, producing, and sharing media history. Elana Levine describes her current book project, a history of the TV soap opera, and how she has used DEVONthink and other digital tools to help organize thousands of hours of recorded broadcasts and an abundance of secondary sources that span

nearly seven decades. Next, Cynthia B. Meyers reflects on her experience using Tumblr to share images from her research into the history of the American advertising industry. Meyers speaks to the tension between wanting to share historical artifacts with a broad public and the growing realization that Tumblr, as a distribution platform, strips these images of their historical context and elicits sexist responses from some users. In contrast to Tumblr, Media Ecology Project (MEP) seeks to broaden access to historical media collections and improve our sense of context in the process. Mark Williams reports on the ongoing development of MEP and its effort to unite scholars, archivists, and software developers around a shared goal. Finally, in the last chapter, Eric Hoyt reflects on the three activities that consume his scholarly work time—curating, coding, and writing. Hoyt argues that we should recognize that these activities involve different processes and generate different products, and we should avoid assuming that any one of them is inherently more valuable or legitimate than the others.

As film and media history research moves forward in the digital age, scholars need to consider what new skills and tools matter most and what existing skills need to be reinvigorated. The contributors to this collection call attention to a range of valuable competencies and tools. Kevin L. Ferguson and Tony Tran, for instance, both use algorithms within their analytical work. Readers seeking to reproduce those forms of analysis will need to spend time becoming comfortable running scripts from the command line. But for researchers who might be intimidated by digital methods, it is worth pointing out that most of the methods described in the book do not require any prior knowledge of computer programming. Laura Horak's and Elana Levine's chapters, for example, describe how scholars can use existing geospatial and research software in their work.

What ultimately unites all the chapters in this book is the recognition that learning new tools requires that we reflect upon

their design. Toward this end, Charles R. Acland advocates for "low-tech" digital tools and cautions against embracing sophisticated algorithms simply because they may be available. Similarly, Miriam Posner, Derek Long, and Deb Verhoeven all reflect upon the assumptions programmed into the databases with which they work. We need to recognize, therefore, that the skill of interrogating software and databases in an informed way is as valuable as the technical mastery of those same programs.

The actuality of our digital age has made critical engagement with digital research methods an essential part of scholarship. While media history is our focus, the issues and illustrations presented in this volume will no doubt speak to a number of other research domains in the humanities. We invite readers to examine and contest the approaches and experiments that appear in the pages that follow. Far from advocating for a unidimensional orthodoxy of digital research, *The Arclight Guidebook* captures the variety of scholarly innovations and hesitations that constitute our scene, research efforts that, in the end, represent the vibrancy of media history today.

ENDNOTES

1 Steven E. Jones, *The Emergence of the Digital Humanities* (New York: Routledge, 2013).

2 Matthew K. Gold, ed., *Debates in the Digital Humanities* (Minneapolis: University of Minnesota Press, 2012).

3 Geoffrey Rockwell et. al., "The Design of an International Social Media Event: A Day in the Life of the Digital Humanities," *Digital Humanities Quarterly 6, no. 2 (2012): np.* http://www.digitalhumanities.org/dhq/vol/6/2/000123/000123.html. This year (2016), the day of DH was held on April 8.

4 N. Katherine Hayles, *How We Think: Digital Media and Contemporary Technogenesis.* Chicago: University of Chicago Press, 2012.

5 Kevin L. Ferguson, "Volumetric Cinema," *[in]Transition* 2, no. 1 (2015), accessed February 19, 2016. http://mediacommons.futureofthe-

book.org/intransition/2015/03/10/volumetric-cinema. Catherine Grant, ed., *[in]Transition* 2, no. 2 (2015), accessed February 27, 2016. http://media-commons.futureofthebook.org/intransition/theme-week/2015/21/journal-videographic-film-moving-image-studies-22-2015.

6 Michael Ross, Manfred Grauer, and Bernd Freisleben, eds., *Digital Tools in Media Studies: Analysis and Research: An Overview* (Bielefeld and New Brunswick: Transcript Verlag, 2009).

7 Tara McPherson, ed., "In Focus: Digital Scholarship and Pedagogy," *Cinema Journal* 48, no. 2 (Winter 2009): 119–60. It is worth noting that several other pedagogy-oriented collections about digital media have been published as well.

8 Catherine Grant, ed., "Film and Moving Image Studies Re-Born Digital?" *FRAMES* 1 (2012), accessed February 19, 2016. http://framescinemajournal.com/?issue=issue1.

9 Jennifer Holt and Alisa Perren, eds., *Media Industries: History, Theory, and Method* (Malden: Wiley-Blackwell, 2009); Charles R. Acland, *Screen Traffic: Movies, Multiplexes and Global Culture* (Durham: Duke University Press, 2003).

10 Charles R. Acland and Haidee Wasson, eds., *Useful Cinema* (Durham: Duke University Press, 2011); Devin Orgeron, Marsha Orgeron, and Dan Streible, *Learning with the Lights Off: Educational Film in the Classroom* (New York: Oxford University Press, 2011).

11 Jon Lewis and Eric Smoodin, eds., *Looking Past the Screen: Case Studies in American Film History and Method* (Durham: Duke University Press, 2007); Richard Maltby, Daniel Biltereyst, and Philippe Meers, eds., *Explorations in New Cinema History: Approaches and Case Studies* (Malden: Wiley-Blackwell, 2011).

12 Fred Gibbs, "Critical Discourse in Digital Humanities," *Journal of the Digital Humanities* 1, no. 1 (2011). http://journalofdigitalhumanities.org/1-1/critical-discourse-in-digital-humanities-by-fred-gibbs/.

13 Benjamin M. Schmidt, "Words Alone: Dismantling Topic Models in the Humanities," *Journal of Digital Humanities* 2.1 (Winter 2012). http://journalofdigitalhumanities.org/2-1/words-alone-by-benjamin-m-schmidt/. Benjamin M. Schmidt, "When you have a MALLET, everything looks like a nail," *Sapping Attention* [Blog], November 2, 2012. http://sappingatten-

tion.blogspot.com/2012/11/when-you-have-mallet-everything-looks.html.

14 Stephen Ramsay, *Reading Machines: Toward an Algorithmic Criticism* (Urbana: University of Illinois Press, 2011).

15 Matthew L. Jockers, *Macroanalysis: Digital Methods and Literary History* (Urbana: University of Illinois Press, 2013). Also, "Beyond Search" served as the title of a 2006–09 workshop series—which eventually became the Stanford University Literary Lab in 2010—helmed by Matthew Jockers.

16 Ted Underwood, "Theorizing Research Practices We Forgot to Theorize Twenty Years Ago," *Representations* 127 (2014): 64.

17 Fred Gibbs and Trevor Owens, "Building Better Digital Humanities Tools: Toward broader audiences and user-centered designs," *Digital Humanities Quarterly* 6, no. 2 (2012), accessed February 12, 2016. http://www.digitalhumanities.org/dhq/vol/6/2/000136/000136.html.

18 Currently in its second phase, Ngram Viewer has increased its corpus to eight million books, using the same procedures for identifying texts for inclusion. See Jean-Baptiste Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books," *Science* 331, no. 6014 (January 2011) and Yuri Lin et al., "Syntactic Annotations for the Google Books Ngram Corpus," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics* (Jeju, Republic of Korea: Association for Computational Linguistics, 2012).

19 Jean-Baptiste Michel et al., "Supporting online material for Quantitative Analysis of Culture Using Millions of Digitized Books," *Science Express* (March 2011), accessed February 19, 2016. doi:10.1126/science.1199644.

20 Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *American Archivist* 68 (Fall/Winter 2005).

BIBLIOGRAPHY

Acland, Charles R. *Screen Traffic: Movies, Multiplexes and Global Culture*. Durham: Duke University Press, 2003.

Acland, Charles R. and Haidee Wasson, eds. *Useful Cinema*. Durham: Duke University Press, 2011.

Ferguson, Kevin L. “Volumetric Cinema.” *[in]Transition* (2015). Accessed February 19, 2016. http://mediacommons.futureofthebook.org/in-transition/2015/03/10/volumetric-cinema.

Gibbs, Fred. “Critical Discourse in Digital Humanities.” *Journal of the Digital Humanities* 1, no. 1 (2011). http://journalofdigitalhumanities.org/1-1/critical-discourse-in-digital-humanities-by-fred-gibbs/.

Gibbs, Fred and Trevor Owens. “Building Better Digital Humanities Tools: Toward Broader Audiences and User-Centered Designs.” *Digital Humanities Quarterly* 6, no. 2 (2012). Accessed February 12, 2016. http://www.digitalhumanities.org/dhq/vol/6/2/000136/000136.html.

Gold, Matthew K., ed. *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, 2012.

Grant, Catherine, ed. “Film and Moving Image Studies Re-Born Digital?” *FRAMES* 1 (2012). Accessed February 19, 2016. http://framescinema-journal.com/?issue=issue1.

Greene, Mark A. and Dennis Meissner. “More Product, Less Process: Revamping Traditional Archival Processing.” *American Archivist* 68 (Fall/Winter 2005): 208–63.

Hayles, N. Katherine. *How We Think: Digital Media and Contemporary Technogenesis.* Chicago: University of Chicago Press, 2012.

Holt, Jennifer and Alisa Perren, eds. *Media Industries: History, Theory, and Method.* Malden: Wiley-Blackwell, 2009.

Hoyt, Eric, Derek Long, Anthony Tran, and Kit Hughes. “*Variety*’s Transformations: Digitizing and Analyzing the First Forty Years of the Canonical Trade Paper.” *Film History* 27, no. 4 (2015): 75–106.

Hoyt, Eric, Kit Hughes, Derek Long, Kevin Ponto, and Anthony Tran. “Scaled Entity Search: A Method for Media Historiography and Response to Critiques of Big Humanities Data Research.” *Proceedings of IEEE Conference on Big Data* (2014): 51–59. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7004453.

Hughes, Kit, Eric Hoyt, Derek Long, Kevin Ponto, and Tony Tran. “Hacking Radio History’s Data: Station Call Letter, Digitized Magazines, and Scaled Entity Search.” *Media Industries Journal* 2, no. 2 (2015). http://www.mediaindustriesjournal.org/index.php/mij/article/view/128/182.

Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History*. Urbana: University of Illinois Press, 2013.

Jones, Steven E. *The Emergence of the Digital Humanities*. New York: Routledge, 2013.

Lewis, Jon and Eric Smoodin, eds. *Looking Past the Screen: Case Studies in American Film History and Method*. Durham: Duke University Press, 2007.

Lin, Yuri, Jean-Baptiste Michel, Erez Lieberman Aiden, Jon Orwant, Will Brockman, and Slav Petrov. "Syntactic Annotations for the Google Books Ngram Corpus." In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*. (Jeju, Republic of Korea: Association for Computational Linguistics, 2012): 169–74.

Long, Derek, Eric Hoyt, Anthony Tran, Kevin Ponto, and Kit Hughes. "Who's Trending in 1910s American Cinema? Exploring ECHO and MHDL at Scale with Arclight," *The Moving Image* (forthcoming).

Maltby, Richard, Daniel Biltereyst, and Philippe Meers, eds. *Explorations in New Cinema History: Approaches and Case Studies*. Malden: Wiley-Blackwell, 2011.

McPherson, Tara, ed. "In Focus: Digital Scholarship and Pedagogy." *Cinema Journal* 48, no. 2 (Winter 2009): 119–60.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, the Google Books Team, Joseph P. Pickett, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. "Supporting Online Material for Quantitative Analysis of Culture Using Millions of Digitized Books." *Science Express* (March 2011). Accessed February 19, 2016. doi:10.1126/science.1199644.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, the Google Books Team, Joseph P. Pickett, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331, no. 6014 (January 2011): 176–82.

Orgeron, Devin, Marsha Orgeron, and Dan Streible. *Learning with the Lights Off: Educational Film in the Classroom*. New York: Oxford

University Press, 2011.
Ramsay, Stephen. *Reading Machines: Toward an Algorithmic Criticism*. Urbana: University of Illinois Press, 2011.
Rockwell, Geoffrey, Peter Organisciak, Megan Meredith-Lobay, Kamal Ranaweera, Stan Ruecker, and Julianne Nyhan. "The Design of an International Social Media Event: A Day in the Life of the Digital Humanities." *Digital Humanities Quarterly 6, no. 2 (2012): np.* http://www.digitalhumanities.org/dhq/vol/6/2/000123/000123.html
Ross, Michael, Manfred Grauer, and Bernd Freisleben, eds. *Digital Tools in Media Studies: Analysis and Research: An Overview*. Bielefeld and New Brunswick: Transcript Verlag, 2009.
Underwood, Ted. "Theorizing Research Practices We Forgot to Theorize Twenty Years Ago." *Representations* 127 (2014): 64–72.

PART I: **SEARCHING AND MAPPING**

THE QUICK SEARCH AND SLOW SCHOLARSHIP: RESEARCHING FILM FORMATS

Haidee Wasson

For historians, time is a complicated matter. We talk about periodization, which is essential, but we also talk about the many registers of historical time: the long durée, event time, epochal time, messianic time. When discussing the practice of research in the context of a rich and diverse set of digital tools, time is also part of the dynamic process that shapes all phases of research. With an enormous volume of digitized materials increasingly available to us, some of the seemingly pedestrian concerns of yesterday's paper-bound historian (painstakingly turning each page of a manuscript, a newspaper, a policy document) have migrated to technical expediency fueled by search terms and algorithms. These digital means allow a previously unimaginable volume of texts to be searched in the amount of time it takes to blink and maybe to stroke a key or two. It appears that what used to take years, can now take nanoseconds. Yet it would be foolhardy to mistake the truly humbling power of the quick search as a replacement for the equally important long-haul of slow scholarship. Looking back, and working toward meaningful and engaged theses involving the past, requires the quick *and* the slow, the aggregate as well as the focused view. Most importantly, writing history still requires good questions that help us navigate the ever-growing body of accessible, searchable stuff.

What follows are some observations made from the media historian's shop floor about changes to scholarly process in light of the recent development of digital research tools. These observations are made in part based on insights generated while working on my current project investigating the history of portable film projectors. The geography of this project is largely American, focusing on the period of 1939 to 1959. This is a cultural history

of technology, one that disarticulates what we tend to think of as the cinematic apparatus by focusing on one constitutive component of it: projectors. This disarticulation has allowed me to chart a distinct technological substrate, a parallel kind of cinema, one that was often linked to display and performance scenarios and embedded in expanded institutional and technological ecosystems not usually included in histories of cinema. The simplest and bluntest goal of my project is to map a particular film format that fundamentally transformed the conditions in which films could be seen. These distinct and enduring viewing conditions served as by far the most common mode by which celluloid was seen and its sounds heard from midcentury until the rise of video.[1] This sea of connected playback devices thus handily displaces the prominence of the movie theater as the historically situated and *de facto* site of cinema.

This project takes its cue from others working on film exhibition, multi-sited, or so-called nontheatrical spaces, and what we assume are minor or "orphaned" genres.[2] It is also shaped by the work of media historiographers seeking to open up and trouble concepts of media that transcend time and reduce the complexity and variation of media-in-history.[3] When it comes to cinema-in-history, the conceptual basis of the project rejects assumptions about a singular or persistently coherent "cinematic apparatus," asserting that cinema has long been iterative during not just its early or late periods but throughout the twentieth century.

Making the claim that small portable film projectors were by far the dominant site for cinema as an everyday technology and experience is somewhat anathema to the ways film histories have largely been written. The prominence of the movie theater in the historical imagination of cinema is something of a truism, and often a precious and magical one.[4] I often feel compelled to begin papers on this sea of machines that I am examining with an apology. Yet, the numbers speak volumes. Piecing together data from

the US Bureau of the Census and reports from the industry, portable projectors outnumbered movie theaters in 1959 by a factor of 263:1. And the numbers simply grow, such that by 1969 portable projectors outnumbered movie theaters by a factor of 632:1 and by 1979, by roughly 1000:1.[5] The contrast of the movie theater completely overshadowed by small projectors creates a challenge of scale, one created for me less by big data but by the big US census, which for two hundred years has been measuring—by law—all manner of data, including, at midcentury, the sale of film projectors.

To be sure, these numbers illustrate some things and not others. Many questions remain. First, it must be said that movie theaters were not only ever one thing. They also have complex histories as hosts for polyglot audiences, multimedial performances, varied programming strategies, and sometimes local, community functions. Yet, their status as built environments with a common link to a particular kind of professionalized apparatus and to commercial programming and distribution practices, distinguish them from portable machines that were designed to be moved, possibly with each operation. That said, these proliferating portable machines are also characterized by significant variation. They used different film gauges or film width; early on they were 9mm and 28mm, but from the early 1920s forward, the most common gauges were 8mm and 16mm with occasional uses of 35mm film, the standard professional gauge which was largely but not exclusively tethered to theaters. Some portable projectors were what we might today call "smart machines" offering many features that allowed for a great degree of control over the film performance. Knobs, buttons, and levers allowed a user to slow down, reverse, or stop the image; to make the images smaller or bigger; to brighten or dim projections; to turn the volume up or down. Such machines allowed for considerable control over all vectors of the projected, amplified form. Other machines offered a more perfunctory set of possibilities, functioning as blunt playback

devices with limited range. Large institutions and individuals used portable projectors differently in public, semipublic, and private ve-nues. These devices were integrated into spaces differently, sometimes displayed prominently, shuffled in on a cart, or hidden in walls, furniture, and behind curtains. These projectors played to large occasional audiences and small recurring ones. Sometimes they were not articulated to audiences at all but simply turned on, visible to people that moved passed them on a busy sidewalk or in a train station.

In short, numbers indicating millions of projectors in use throughout the United States did not help on their own to parse the specific capacities of any given group of devices or the diverse frequencies and methods of their use. Greater specificity required thumbing through trade journals looking for advertisements, company records detailing product development, and design and technical discourses as recorded in meeting minutes, user manuals, technical protocols, and product reviews.[6] Yet, it must be said that the numbers make a powerful point about where cinema *happened* throughout the twentieth century. And, at least by midcentury, *it is not what most of us suppose*, making projectability—that is, the widespread capacity to project—a salient concept for histories of cinema. The many possible forms that this capacity to project took or may have taken constitute the specific contours and the telling insights that something like a mass of statistics about millions of projectors might mean for film and media history. This unique and previously uncharted capacity now requires the slower, steady work of embedding these machines within the complex social, cultural, and political systems, as well as the people, groups, audiences, and institutions that help us to understand how these technologies mattered in history and beyond.

This big clear picture has forced me to find more manageable points of entry, to distill a better understanding of the differences that conveniently add up to the sameness implied by millions

of devices designated by the term "portable projector." So, my project works from this scale of projector-millions and identifies particular nodal points of focused investigation that help to tell important elements of this larger story. I am working to link particular and specific machines to aesthetic, experiential, cultural, political, sonic, and audile imperatives. What were the media ecologies, cognate display practices, related design imperatives, and instrumental functions that portable projectors were linked to? And thus what were the practices and aesthetics that were enabled by those projectors? I make these links by looking at institutions important to, but not conventionally understood as germane to, the history of cinema: the fair, the military, industrial design, the home, the museum. Drilling down into this range of institutions has involved accessing a broad range of sources (film and photographic trade literature, the *Journal of the Society of Motion Picture Engineers*, technical manuals, sales catalogues and pamphlets, government documents, military design protocols, industrial design documents, intra-organizational records, museum bulletins, newspaper articles, and popular literature). Along the way, I am working to identify exceptional and specialized devices, powerful and prosaic performance scenarios, and the distinct film types that these uses of technology yielded (among them film memos, munitions test films, poster films, product demos, point-of-sale films, industrial spectacle, and ambient cinemas). Ideas about spectatorship and viewers change considerably as some of the films shown on these machines circulated conventionally by way of national or regional distributors to local exhibitors, but many other films sat on shelves and had a kind of reference function. They accompanied products and served as operating manuals or they were for intra-organizational uses, made for audiences of one or two. Some films were made with the clear sense that they would hardly or perhaps never be seen at all, functioning rather abstractly as records with an unknown use in the future.[7] So, within the parameters of this project that is still admittedly in formation, I want to weave together some observations about

the changing research landscape and how this project has been shaped by an expanding digital toolkit. The crucial activity and idea of "searching" subtends each of the following three interjections.

PARSING THE DIGITAL: SMALL TOOLS

We know that what we call the digital has many grooves, contours, textures, speeds, and uneven traffic flows. For my work so far, this insight helps me to make sense of my particular relationship to digital research, best characterized as a practice of using select tools and maybe even what we might call select *small* tools. To be sure, these tools range in scale, importance, function, and frequency of use. They are not just small. I use the bigger ones, as well. They include, of course, full-text searchable professional journals, trade journals, newspapers, and periodical literature. They also include web-based platforms for distribution of moving images and sounds, such as the Internet Archive, the Library of Congress, YouTube, and eBay. With smaller and more specialized sites like the Canadian Education, Sponsored, and Industrial Film (CESIF) database, the US Department of Energy, and a wide range of fan and collector sites, there exists an eclectic but determined treasure trove of moving images and sounds that used to be called "ephemeral" yet now have second lives. It is easy to forget how much of the history of film is based on the condition of scarcity. This in and of itself has changed significantly with digital collections and web-based viewing interfaces where now online moving images are offered up in an unprecedented and resplendent diversity.

Many of the small digital tools that I use entail modes of converting what was once paper to digital formats, and this has radically transformed what these documents are and how they function as part of my own research and writing practice. Many others have talked about this with regards to newspapers.[8] Documents become accessible as well as highly moveable and searchable not just once

but over and over again. This insight extends to innumerable kinds of documents: finding aids, bibliographies, the holdings of individual libraries and special collections. Even those tools that by some measures might seem banal—the "finder" function on my desktop—have helped me to keep track of the digital equivalent of cocktail napkin notes and to integrate my sporadic fits of clarity into chapter drafts. The finder button helps me to connect notes that I made years ago under a pressing writing deadline to the present, and it assists in scouring documents once discarded for a second or third consideration. Add to this file protocols like PDF, JPEG, WAV, scanning apps like CAMSCAN, which turns my iPhone into a portable document scanner, and a small army of procurement devices like digital cameras, smart phones, and tablets that help us to copy, upload, store, rearrange, and transfer images, texts, and sound files. No question, I still use a photocopier. But I also use cloud-based data-storage services and portable hard drives and thumb drives. These tools allow me to carry these files, to store, back up, access, and share them, to ultimately transform them into organized thought, and to present them to students and colleagues. Small digital tools have made organizing an unseemly volume of material something of a constitutive chore, but they have also meant that organizing and reorganizing data, files, and emails becomes a core practice of research itself, where we are constantly re-labelling and grouping vast amounts of data within our own project-based collections. At its best, this activity is highly generative, regularizing and amplifying a longstanding recombinant and recursive gesture into the process. At its worst, it is distracting, inefficient busy-work. But analog and digital processes have always entailed both the best and the worst tendencies of its media and its methods. It is best to be honest and clear about both. Last, it should be said that big data and small tools can and probably should work together, transforming research fundamentally in ways both dramatic and quietly so.

PARSING THE DIGITAL: MULTIMODAL RESEARCH

Where might the digital end and all other modes of engaging with a world of evidence and modes of analysis begin? I use a range of research methods that are highly integrated with my digital tools, but are not reducible to them: their points of contact and mutual entanglement are often invisible or seamless to me. I still make post-it notes, scribble on typed documents, scrawl on note pads and the backs of envelopes. When possible, I watch films on celluloid. I still talk to humans—scholars, archivists, technicians, projectionists, and curators—to draw on their expertise; they sometimes expand my mess and sometimes help me make sense of my research sprawl. I also touch objects, hold them, and sometimes when nobody is looking I actually play and fiddle with them. For this project, I look at and touch collapsible screens and gun cameras. I try to move or carry desktop and suitcase projectors, machines that were called "light weight" and consoles that were not. Nothing reminds you about the relative meaning of portability faster than picking up a 75-pound portable projector. The things in our research are not reducible to their physicality but we also learn through touching, using, breaking, fixing, and photographing them. It seems to me that working in multiple modes and developing dynamic research and writing strategies will help serve us best as we continue to articulate our searches to algorithms and develop algorithms that respond to our research. Ultimately, algorithmically derived data will be most meaningful when put in dialogue with a range of other objects, evidence, and methods.

PARSING THE DIGITAL: SEARCHING

Lastly, searching transcends format and medium and algorithm and is fundamental to scholarship. As researchers we search. We search for evidence that confirms our thesis and hopefully for evidence that does not. We also search because we are unsure of what we will find. We search to search again and to identify particular places to go, things to examine, and vectors to travel.

Searching is a diagnostic process as much as a cure. And, as Ted Underwood reminds us, we search for things that are named differently depending on time and contexts.[9] As scholars we use concepts to frame a project, which then shape our searches. In my case, I have framed my project around "portability" and "projection" but that doesn't mean that what I need to investigate can be found by recourse to those terms alone. With a resource like the Media History Digital Library, I can easily see how a term like portability is relative both synchronically and diachronically; it means different things depending on how and where it is being used. What portability was in 1918 within the American film industry was inextricably linked to flammability. That is, if a projector was considered portable it was first and foremost not incendiary. Fifteen years later the American military, which regularly reported to the American film industry on its innovations and needs, linked portability more to a quality that might best be described as ruggedness. Could it be dropped ten times and still work? Beyond printed film matter, the term portable was widely applied to all manner of media (radios, floodlights, theater stages). To carry out a project like this using keyword searches across large and varied materials, I need to use a range of terms and then do contextual digging to assess relevance. "Portable projector" was a term that was occasionally used. But often this fact of portability and projection was not always noted upon and it has to be inferred by virtue of location (basements, union halls, airfields) or institutions (schools, governments) or perhaps by use of technical terms such as gauge (16mm or 8mm). Even particular kinds of films can often be reasonably used to index the presence of a portable projector (educational, training, sponsored films). So, I also search familiar terms like small or substandard, miniature, amateur, nonprofessional, and travelling. Terms that I have rejected conceptually as a way to organize the project itself (nontheatrical) are used in the searching process simply because they were highly used terms at certain historical moments, indexing events and phenomena relevant to the broader project. But,

digital searching has also made possible a kind of playful searching that would have been highly formidable and really quite impossible before the sizable databases we have now. So, I also search for terms like brightness, noise, hiss, broken, dirty, repair, fix, interruption, damage, fuzzy, blurry, heavy, film loops, viewing boxes, projection cones, cinemobiles, film trucks, projection tents, and airstrip shows. We can now be more nimble and playful with language as we scan enormous bodies of evidence and use its beguiling charms carefully but also obliquely and creatively. I can take these terms and apply them not just to film journals but to newspapers and magazines that are local, regional, national, and international, and to photography, theater, design, business, governmental, and industrial literatures, each of which might have both highly specialized and generalist versions. These searches may yield little but frequently they offer up a nugget, a curiosity, a question that deserves further and more focused thought.

Lastly, we must ask good questions, which are themselves like compasses whose points of orientation can be recalibrated as we travel down the research road. Searching can be a straight-line drive down a highway, but it should just as often be an experimental and exploratory wandering that includes a kind of consequence-free play of associations and lateral leaping. We did not need the digital to think of searching this way. But it has surely enabled this process, making more feasible cross-disciplinary, exploratory approaches to our questions across a wider and wider research landscape.

The project I have outlined here requires materials that only exist in single places and that exist everywhere. It requires small tools and big. Ultimately as media historians and scholars, the digital and the algorithmic must be integral but partial elements of a fuller process: constituting elements of the workstations on our shop floor that involve—from beginning to end—asking important and generative questions that guide the process of searching, gath-

ering, analyzing, testing, and ultimately presenting our work. This process entails calibrating and recalibrating questions in a kind of recursive gesture, questions that are scaled to the kind of evidence we have, and adjusting them as we learn that they are too big, too general, too narrow, and (gasp!) unanswerable. Ultimately, searching allows us to mix a degree of open-ended experiment with a rigorous eye to identifying supportable claims, made possible by an expanding body of evidence that responds to questions that matter. Alongside the quick and powerful search, we must also recall the long view to slow research.

ENDNOTES

1 For more detailed documentation see Haidee Wasson, "The Protocols of Portability," *Film History* 25, no. 1-2 (2013): 236–47; and Haidee Wasson, "Suitcase Cinema," *Cinema Journal* 51, no. 2 (Winter 2012): 148–52.

2 See for instance Lee Grieveson and Colin MacCabe, eds., *Empire and Film* (London: British Film Institute, 2011); Devin Orgeron, Marsha Orgeron, and Dan Streible, eds., *Learning with the Lights Off: Educational Film in the United States* (New York: Oxford University Press, 2012); Vinzenz Hediger and Patrick Vonderau, eds., *Films that Work: Industrial Film and the Productivity of Media* (Amsterdam: Amsterdam University Press, 2009); and Charles R. Acland and Haidee Wasson, eds., *Useful Cinema* (Durham: Duke University Press, 2011).

3 See for instance Jonathan Sterne, *MP3: The Meaning of a Format* (Durham: Duke University Press, 2012) and Lisa Gitelman, *Always Already New: Media, History, and the Data of Culture* (Cambridge: MIT Press, 2006).

4 For this foundational assumption in theories of "the apparatus" and the determined camera- and celluloid-centric ideas about the apparatus, see Jean-Louis Baudry, "The Ideological Effects of the Basic Cinematographic Apparatus," in *Narrative, Apparatus, Ideology: A Film Theory Reader*, ed. Phil Rosen (New York: Columbia University Press, 1986), 286–98; and Teresa De Lauretis and Stephen Heath, eds., *The Cinematic Apparatus* (New York: St. Martin's Press, 1980). For important historical

work see Richard Koszarski, *An Evening's Entertainment: The Age of the Silent Feature Picture, 1915–1928* (Berkeley: University of California Press, 1990). For a more recent work that looks back in interesting ways but also nostalgically see Gabriele Pedullà, *In Broad Daylight*, trans. Patricia Gaborik (London: Verso, 2012).

5 Statistics on projectors were published regularly in the annual reports entitled *The Wolfman Report on the Photographic Industry in the United States*, compiled by Augustus Wolfman. These reports began to appear in the mid-1950s in annual issues of *Photo Era* magazine and during later years were published by *Modern Photography* and *Popular Photography* as stand-alone publications until the early 1990s. Wolfman compiled statistics from the US Bureau of the Census as well as published studies issued by elements of the audio-visual industry. He died in 1974, though reports continued to be issued using his name.

6 Key collections include the Victor Animatograph Collection at the University of Iowa, the Ellis Collection at Duke University, which holds a treasure trove of material on Kodak, and the John W. Hartman Center, also at Duke University, which holds the J. Walter Thompson Company Archives. The New York Public Library holds innumerable resources in its business library and its Manuscripts and Archives Division, including records related to the 1939 New York World's Fair. Issues of the *Journal of the Society of Motion Picture Engineers*, which are available through the Media History Digital Library, have also been invaluable.

7 James Paasche has examined, for instance, the case of the Department of Army Special Photographic Office, a group of photographers and filmmakers within the US Army who worked under the mandate to document the army's efforts during the Cold War. Formed in 1962, the unit exposed hundreds of thousands of feet of film as a record of army activity and the everyday work and life of enlisted men. Most of this footage was never processed or edited into what we might call finished films. See James Paasche, "Shots Made Around the World," in *Cinema's Military Industrial Complex*, eds. Lee Grieveson and Haidee Wasson, (Berkeley: University of California Press, forthcoming).

8 See Gregory A. Waller in this volume and Richard Abel, "The Pleasures and Perils of Big Data in Digitized Newspapers," *Film History* 25, no.

1-2 (2013): 1.

9 Ted Underwood, "Theorizing Research Practices We Forgot to Theorize Twenty Years Ago," *Representations* 127, no. 1 (Summer 2014): 64–72.

BIBLIOGRAPHY

Abel, Richard. "The Pleasures and Perils of Big Data in Digitized Newspapers." *Film History* 25, no. 1-2 (2013): 1–10.

Acland, Charles R. and Haidee Wasson, eds. *Useful Cinema*. Durham: Duke University Press, 2011.

Baudry, Jean-Louis. "The Ideological Effects of the Basic Cinematographic Apparatus." In *Narrative, Apparatus, Ideology: A Film Theory Reader*, edited by Phil Rosen, 286–98. New York: Columbia University Press, 1986.

Gitelman, Lisa. *Always Already New: Media, History, and the Data of Culture*. Cambridge: MIT Press, 2006.

Grieveson, Lee and Colin MacCabe, eds. *Empire and Film*. London: British Film Institute, 2011.

Hediger, Vinzenz and Patrick Vonderau, eds. *Films that Work: Industrial Film and the Productivity of Media*. Amsterdam: Amsterdam University Press, 2009.

Koszarski, Richard. *An Evening's Entertainment: The Age of the Silent Feature Picture, 1915–1928*. Berkeley: University of California Press, 1990.

Lauretis, Teresa De and Stephen Heath, eds. *The Cinematic Apparatus*. New York: St. Martin's Press, 1980.

Orgeron, Devin, Marsha Orgeron, and Dan Streible, eds. *Learning with the Lights Off: Educational Film in the United States*. New York: Oxford University Press, 2012.

Paasche, James. "Shots Made Around the World." In *Cinema's Military Industrial Complex*, edited by Lee Grieveson and Haidee Wasson. Berkeley: University of California Press, forthcoming.

Pedullà, Gabriele. *In Broad Daylight*. Translated by Patricia Gaborik. London: Verso, 2012.

Sterne, Jonathan. *MP3: The Meaning of a Format*. Durham: Duke Univer-

sity Press, 2012.
Underwood, Ted. “Theorizing Research Practices We Forgot to Theorize Twenty Years Ago.” *Representations* 127, no. 1 (Summer 2014): 64–72.
Waller, Gregory A. “Search and Re-search: Digital Print Archives and the History of Multi-Sited Cinema.” In *The Arclight Guidebook to Media History and the Digital Humanities,* edited by Charles R. Acland and Eric Hoyt (Falmer: REFRAME/Project Arclight, 2016).
Wasson, Haidee. “Suitcase Cinema.” *Cinema Journal* 51, no. 2 (Winter 2012): 148–52.
———. “The Protocols of Portability.” *Film History* 25, no. 1-2 (2013): 236–47.

SEARCH AND RE-SEARCH: DIGITAL PRINT ARCHIVES AND THE HISTORY OF MULTI-SITED CINEMA

Gregory A. Waller

There is no question that media historians should be aware of the scholarly possibilities of data visualization and topic modeling, as Eric Hoyt has argued, and explore the opportunities for searching vast quantities of digitized information with tools like Arclight.[1] In this essay, however, I will discuss a messier and—for the researcher—more labor intensive engagement with digital archives of print material. My test case is a project on the history of nontheatrical cinema, an area of media history that poses unique challenges beyond being largely overlooked by film historians. This project is less concerned with big data than with the specificities and minutiae of digitized discourse explored at the micro level.

I begin with a fairly straightforward question: what was nontheatrical cinema in 1915? That year saw *The Birth of a Nation,* Charlie Chaplin's meteoric rise, the opening of Universal City, and the US Supreme Court's *Mutual* decision, which denied First Amendment protection to the producers of motion pictures—all highly visible signs that the American film industry was either fully or well on its way to being consolidated, or institutionalized, borrowing André Gaudreault's formulation.[2] To put the question another way, what was nontheatrical cinema several years before the term itself even came into use, before trade magazines like *Moving Picture Age* (1919–22) and *Educational Film* (1922–62) began publication, and almost two decades before 16mm became widely adopted in the United States as the format of choice for production and exhibition outside the commercial film industry? How can digital archives of print material help us begin to answer this question?

My use of these archives, as will soon become apparent, is very much conditioned by my previous research on the history of film

exhibition, both theatrical and nontheatrical. In seeking information about the nontheatrical in 1915, I assume, for example, that in the years before the arrival of commercial television: (1) the screening of films outside of commercial movie theaters was neither a monolithic nor unchanging phenomenon but varied from place to place and time to time in several ways, including how this type of exhibition was put into practice, how it was promoted and marketed, and how it was imagined and discussed by advocates, audiences, and producers; (2) the nontheatrical was always positioned—implicitly or explicitly—in relation to something understood as the theatrical, and yet at the same time it was never neatly homologous with the educational; and (3) that nontheatrical cinema might more accurately be described as what I have called *multi-sited cinema* as a way of acknowledging and foregrounding how, where, and by whom it was exhibited.

INSTANTIATION

Digital archival resources offer an unprecedented opportunity for undertaking what is the necessary first step for a history of multi-sited cinema—*instantiation.* I do not borrow this term from computer science, but take instantiation to mean finding any evidence of a screening of a film outside of a theater whose primary product was the movies regularly delivered to paying audiences, day in and day out. (Keeping in mind that in the 1910s, moving picture theaters sometimes showed advertising films, could be used for a variety of events, and could even be rented out for "free" screenings—all conditions that potentially blur the line between the theatrical and the nontheatrical.) Beyond identifying screenings, instantiation can provide information about the frequency and variety of certain exhibition sites and practices, the targeting of specific audiences, the role of sponsorship, and the makeup of programs (including multiple-media programs that relied on lantern slides, live performance, and the spoken word along with motion pictures). Instantiation might help reveal whether or not the 1910s saw the emergence of explicitly nontheatrical films, the

privileging of certain genres, the advent of specialized distribution systems, the widespread practice of government and corporate sponsorship of film exhibition, and the establishment of protocols for how such films were to be used.

If instances of nontheatrical screenings in 1915 were to be discoverable anywhere, it would be in daily and weekly newspapers. Or so I figured based on my previous research on moviegoing and film exhibition during the silent era in Lexington, Kentucky. That project had required a page-by-page, day-by-day reading of microfilm copies of both daily Lexington newspapers between 1895 and 1930, and it was in these newspapers—and not in the trade press like *Moving Picture World*—that I discovered traces of other exhibition sites and occasions apart from the movie theaters on Main Street, including novel outdoor uses of advertising film, public screenings of industrial films at the University of Kentucky, and, most significantly, motion picture programs presented at African American churches and "colored" schools.[3] It is worth noting that it would be impossible for me to replicate this study in 2015 using digital resources, since only the city's Democratic morning newspaper has been digitized. Yet it was the Republican evening paper that contained by far the most richly suggestive bits and pieces about the amusements available for African Americans in segregated Lexington.

Given my sense that newspapers constitute an invaluable primary source for exploring the early history of multi-sited cinema, instantiating the nontheatrical in 1915 seemed to be a task ready-made for digital newspaper archives and an effective search platform. However, the potentially definitive search term, *nontheatrical,* did not yet exist in 1915, and neither, of course, did the term *multi-sited.* What's more, the non-theater was not a specific type of exhibition space but rather an open-ended array of possible sites, particularly once projectors had become somewhat portable, which was the case by the mid-1910s. My solution was to start

broadly and inclusively, by examining the results from individual dates in four newspaper archives, using the search terms: *moving picture, motion picture, film, movies, photoplay,* and *cinema*. The archives searched were the Library of Congress site, Chronicling America: Historic American Newspapers, and three for-profit archives that have some overlap in their holdings: newspapers.com, genealogybank.com, and newspaperarchive.com.[4]

Digital newspaper archives contain what appears to be a massive and ever-growing amount of information. Each of the three commercial archives boasts of having more than a billion pages searchable with more added each month. Yet, however big they are or might become, these archives are and always will be frustratingly incomplete, random, and unrepresentative in their holdings (not to mention the problems caused by optical character recognition). Much more than the Media History Digital Library with its admirably completist aspirations, digital newspaper archives always bear witness to the vagaries of time and chance and to the often now-invisible decisions over the years by various publishers, gatekeepers, collectors, archivists, funding sources, and profit-minded owners. As such, these convenient (if sometimes costly) resources are an object lesson in the historiographical ground rules when writing media history based on surviving print discourse. Even though the accessible, searchable newspaper record is necessarily neither complete nor representative, individual search results are no less interesting and potentially generative, offering leads to be followed into the still vast expanse of digital space.

To begin, I examined two dates during each month of 1915 (one Thursday and one Sunday). Through all the 1915 dates searched, none of the terms—*moving picture, motion picture, film, movies, photoplay, cinema*—appears exclusively or primarily in relation to nontheatrical film. Instances of screenings outside of moving picture theaters are most likely to come up in searches for *mov-*

March 18, 1915	motion picture	moving picture	photo-play	film	cinema	movies
newspapers.com	118	113	64	209	1	62
genealogybank.com	44	44	41	124	2	41
newspaperarchive.com	49	56	46	184	2	62
Chronicling America	41	51	52	97	1	51

Figure 1. Tabulation of results, March 18, 1915, indicating number of pages containing the search term. Note that these online archives sometimes hold the same newspapers and that individual pages might include several uses of the same term.

ing picture and *motion picture*, while *photoplay* and *movies* very rarely appear in connection with the nontheatrical. *Film*—by far the most frequently used of these terms—refers with few exceptions to some aspect of the commercial film industry (individual titles, production and distribution companies, etc.), to state censorship activities, or to the photography business. On occasion, the terms could become somewhat interchangeable in descriptions of nontheatrical exhibition, as when the *Evening Missourian* (Columbia, Missouri) used *moving picture*, *film*, and *'movies'* (sic) in reporting on plans to screen two films at the University of Missouri's auditorium during the school's "Journalism Week."[5] Or when the National Press Club in Washington, DC, hosted the Bolivian minister to the United States who screened promotional films about his country, and the *Washington Post*'s brief article on the screening referred to the footage as *movies*, *films*, and a "series of *moving pictures*," while jokingly dubbing the minister a "*motion picture* impresario."[6]

Sifting through, collating, and organizing the instantiations culled from digital newspaper archives highlights certain nontheatrical sites, practices, and sponsors that were particularly prominent and/or widespread in 1915. For example, the Bolivian films mentioned by Washington newspapers were destined for use in an exhibit at that year's world's fair in San Francisco, the Panama-Pacific International Exposition, which saw the regularly scheduled screening of motion pictures sponsored by manufacturers, government agencies, nations, and American states. Other standouts include certain ambitious advertising campaigns that deployed films, lectures illustrated with moving pictures, and the manifold ways that churches utilized motion pictures. Researching these and other iterations of the nontheatrical is quite literally a matter of re-searching, beginning with specific instantiations and returning again and again to the digital newspaper archives for new searches. It is in this process of re-searching that the particular capabilities and opportunities digital resources afford to the historian of nontheatrical cinema (and perhaps to media historians more broadly) become most strikingly evident.

CONTEXTUALIZING AND BRANCHING OUT FROM THE INSTANCE

I'll focus, for no special reason, on Thursday, March 18, 1915. Newspapers from this date yield a number of intriguing potential starting points for re-search: the Washington Press Club screening of Bolivian films mentioned above; a "film lecture" on Scandinavia given at the Danish Brotherhood Hall in Racine, Wisconsin; baseball instruction using moving pictures at Harvard University; moving pictures included as part of St. Patrick's Day festivities for Irish Americans at a 5,000-seat hall in Chicago; moving pictures and stereopticon views featured daily at the first State Conference and Exhibit on Mental Hygiene in Albany, New York; and films promoting immigration to Minnesota shown by a former congressman to an audience in rural Iowa.[7] As an example of where an instance might lead, I will begin with a nontheatrical screening

at a social gathering of telephone company employees in Charlotte, North Carolina.[8]

On May 18, 1915, the *Charlotte News* reported that the "Telephone Society" of Charlotte, North Carolina, had held a "delightful" social night at the local Young Men's Christian Association (YMCA) featuring three motion pictures, which were part of an "interesting and instructive" program. Presented "under the auspices" of the telephone company's "Safety First" committee, this program included a talk on "the prevention of disease and first aid" by a physician employed by the company. Before and between the three films, there were vocal and piano performances by several of the "women employees" of the company who were specially invited guests for this occasion. Individual films exhibited nontheatrically are typically not identified by title in newspaper notices, but in this case the *Charlotte News* noted the names of the films screened for the Telephone Society: *The Man He Might Have Been, The Crime of Carelessness,* and *The Workman's Lesson*—films that had been provided for the event by a member of the "industrial department" of the YMCA's "international committee."[9] There is no mention by the *Charlotte News* that these three films had initially been released in 1912 or that they had been produced for and widely circulated by the National Association of Manufacturers (NAM).[10]

Needless to say, without access to digitized and searchable newspaper content it would have been virtually impossible for me to become aware that a screening for the Telephone Society at the Charlotte YMCA had taken place on a March evening in 1915. Once a search relying on general terms (like *moving picture* or *film*) has yielded an individual instance like this Telephone Society screening, the next step in gathering information might be a follow-up (or "advanced") search that couples *moving picture* with a date range and additional words or phrases. I will be working through a different re-search strategy, however, that begins with culling

from the individual instance specific details related to what I take to be certain primary variables involved when motion pictures were exhibited outside of theaters from the 1910s through at least the 1930s:

* What were the film or films screened?
* What was on the program, in addition to motion pictures—live performances, lantern slides, speeches, discussion?
* What was the actual location, the screening site?
* Was the screening occasion a unique or a regularly scheduled event?
* Under whose auspices or sponsorship were moving pictures exhibited?
* Who was the targeted audience? Was the audience restricted? Invited? Captive?

There is no guarantee, of course, that a newspaper's reference to a nontheatrical screening will necessarily provide answers to any or all of these questions. Rather than searches that rely on general terms like *program, site, occasion, sponsor,* or *audience,* what is needed are unique searches based on whatever information is provided about the individual instance. In the case of my example from Charlotte, this means another round of searches that I will describe below, searches based on the particular film titles (*The Man He Might Have Been, The Crime of Carelessness,* and *The Workman's Lesson*), the site (YMCA in Charlotte), and the sponsor (telephone society). Relying fully on the availability of and access to searchable digital newspaper archives, this process aims at contextualizing and, equally important, branching out from the *Charlotte News*' account of this single nontheatrical screening.

THE MAN HE MIGHT HAVE BEEN, THE CRIME OF CARELESSNESS, AND THE WORKMAN'S LESSON

Open-ended searches using the titles of the three films shown for the Charlotte Telephone Society make clear that all three films

had been in circulation well before March 1915. The newspaper record for *The Man He Might Have Been* begins in 1913 when this film was booked into a range of theatrical venues, from the College Theater in Seattle, Washington, to the Dixie Theatre in Bryan, Texas.[11] In such cases, this one-reeler was scheduled with three or four other films fitting into a standard type of modular program. If *The Man He Might Have Been* was described at all in advertisements and promotional notices for these screenings, it was billed as an Edison production. The other two films shown for the Telephone Society also were at first circulated theatrically, sometimes booked into the same theater, like the Pastime in Pendleton, Oregon, where *The Workman's Lesson* was screened in July 1912, *The Crime of Carelessness* in January 1913, and *The Man He Might Have Been* in February 1913.[12] At virtually every stage in the process of re-searching and historical contextualization, there are choices to be made about which lines of inquiry to follow. For instance, the information above could lead to searches designed to find evidence of the promotional strategies for Edison releases or the programming of Pendleton's Pastime Theater in 1912–13.

But my interest was specifically in the exhibition history of *The Workman's Lesson, The Crime of Carelessness,* and *The Man He Might Have Been.* From at least December 1913 until 1916, these three films were exhibited in sites, on occasions, and to targeted audiences quite distinct from the everyday practices of the commercial moving picture theater. If searches of digital newspapers can never deliver anything that resembles comprehensive coverage of when and where certain films were shown, instantiation can suggest the parameters of possibility. *The Man He Might Have Been* was shown, for example, by the "employee's safety organization of the Central Hudson Gas and Electric Company" (in Poughkeepsie, New York) to this company's workers, as well as being screened for the attendees of a national Citizenship Convention held at a normal school in Washington, DC. *The Workman's Lesson* figured as part of the "Kansas City Safety Rally," with ac-

companying remarks by officials from two railway companies.[13] Movie theaters could also be enlisted (or perhaps rented) for such screenings; such as when the Nevada Industrial Safety Association presented all three NAM films at the Grand Theater, in Reno, Nevada, or when the Tri-City Manufacturers' Association of Rock Island, Illinois, arranged for them to be shown on different days at different theaters to groups of school children.[14] These last two examples testify to the not infrequent case of the moving picture theater temporarily becoming, in effect, a nontheatrical site.

Once they began a second and far more robust multi-sited life outside of commercial film exhibition, *The Man He Might Have Been*, *The Crime of Carelessness*, and *The Workman's Lesson* were frequently screened together, or sometimes in pairs, and identified as having been provided by or produced by the National Association of Manufacturers. (Of course, the connection to NAM could also have been made readily apparent to the audience by a speaker at any screening of the films.) They were presented, for instance, "under the auspices" of the Bridgeport [Connecticut] Manufacturers' Association, with local factory foremen and superintendents in attendance along with NAM officials.[15] Representatives of the extension division of what was then Iowa State College tried to drum up support for a proposed "circulating library of motion pictures" serving public schools by screening and discussing these films at the high school in Keokuk, Iowa, as well as at the "Mothers' Club" of the First United Evangelical church in Marshalltown, Iowa, with the screening held under "the direction of the City Federation of Women's Clubs."[16] Again, any one of these instances might generate additional searches, concerning, for example, the role of public colleges, the Federation of Women's Clubs, and the United Evangelical church as sponsors of screenings outside of commercial venues.

Tracking the circulation of NAM films turned up several instances where the YMCA figured prominently, either as an exhibition

site or a sponsor—or both. In fact, a year before these films were shown to the Telephone Society, the Charlotte YMCA had already screened them in February 1914 as part of its "educational" outreach efforts, charging no admission and welcoming anyone willing to show up for the event.[17] When the NAM films screened at the YMCA in Wilkes-Barre, Pennsylvania, the sponsor was the local chamber of commerce, which incorporated the films into an "Industrial Betterment" meeting. The Anderson, South Carolina, YMCA itself seems to have served as the sponsor when it took *The Man He Might Have Been* and *The Workman's Lesson* out into the community, screening these films on-site at various mills in the area.[18] Thus, even a preliminary reading and collation of search results for the three films points not only to their multi-year, multi-sited exhibition but also to the relation between theatrical and nontheatrical venues, programs, and targeted audiences as well as to the way that the circulation of individual titles reveals links among sponsors, educational campaigns, public institutions, and venues.

THE CHARLOTTE TELEPHONE SOCIETY

For a second series of searches prompted by the March 1915 exhibition of the NAM films in Charlotte, I focused on the particular occasion, sponsor, and audience for this screening, the local Telephone Society, expanding the temporal field to cover from 1910 to 1916 while initially sticking only to Charlotte newspapers. According to various items in the local press, the Charlotte Telephone Society was formed in November 1914 by the workers of the Southern Bell and American Telephone and Telegraph Companies. It initially had a charter membership of eighty-three, limited to male office employees, with "associated" membership available to male workers for other companies like Western Electric and to men "interested in telephone affairs." The goal of this organization, following the model of similar groups of Bell employees elsewhere, was "promoting broader views and bringing about closer and more intimate relations between [male] employees."[19] The

Charlotte group's first meeting, held on November 17, 1914, in the ballroom of a local hotel, largely consisted of a round-table discussion and a series of vocal and instrumental performances by society members, making the evening something of an amateur talent show.[20] Subsequent monthly meetings featured similar musical performances as well as a lecture or a discussion concerning a topic usually related to the telephone business at the local or state level.[21] A special attraction at the group's 1915 Christmas banquet was a long-distance phone call from Southern Bell's president in Atlanta.[22] Judging from these results, the March 18, 1915, meeting of the Telephone Society thus seems to have been atypical in that it included specially invited female employees of the company, featured motion pictures, and was held at the YMCA. There is no evidence one way or the other of any connection between scheduling motion pictures and inviting female coworkers to the meeting. Did the screening of films help make this gathering, in the words of a local newspaper, a "delightful" social event? Were the films (unlike the musical performances) a way to maintain—even in the midst of delight and sociality—a distinctly instructional focus on workplace safety and individual responsibility?

Branching out beyond Charlotte newspapers for information about telephone societies more generally indicates that similar organizations were active across the United States. While I found no evidence that the same films exhibited in Charlotte were shown at other telephone society meetings, such meetings sometimes did include film exhibition. For example, a moving picture "showing the construction and operation of the Bell transcontinental telephone line" was screened for two telephone societies in 1915. After watching Bell's film about its transcontinental telephone, members of the telephone society of Washington, DC, (composed of employees of the Chesapeake and Potomac Telephone Company) listened to a talk by a manager of the firm that helped construct this line. By way of contrast, in Harrisburg, Pennsylvania, moving pictures of the work that went into this major Bell investment in

telecommunications were scheduled for the conclusion of a meeting, preceded by a Charlie Chaplin movie and vaudeville-style talent show featuring an orchestra, magician, vocal quartet, brass quartet, comic monologist, and buck-and-wing dancer.[23] The difference between these two telephone society screenings—coupled with the example from Charlotte—underscores the significance of programming as a key variable in thinking about how moving pictures were positioned, framed, and deployed outside of the movie theater. While searches using *telephone society* do not indicate that these employee social groups regularly or frequently made use of motion pictures, even a handful of instantiations means that the telephone society stood as one more location for multi-sited cinema, circa 1915. While we can never come close to knowing the full extent of film exhibition outside of movie theaters, mapping the options available for multi-sited cinema remains a necessary step in thinking about the dispersion and practice of the nontheatrical.

THE CHARLOTTE YMCA

Most likely the Charlotte Telephone Society shifted its March 1915 meeting from a hotel to the YMCA because this site offered access to projection equipment and/or to a space that allowed for motion picture exhibition. What role did film exhibition play in the overall outreach strategy of and many activities hosted by this YMCA? That's a question we can begin to answer by sifting through the many mentions of the YMCA in Charlotte's newspapers during 1915. That year the local YMCA gained the most attention from the press for the organization's involvement with youth athletics and the Boy Scouts and for the religious activities it scheduled, including lectures by missionaries, monthly meetings of the Ministerial Association, and regular Sunday church services.[24] But this particular YMCA also functioned as something of a multipurpose, more secular venue by serving as the site for various events, including musical recitals, a banquet for Civil War Veterans on the occasion of Lee-Jackson Day (commemorating the birthdays of

Confederate generals Robert E. Lee and Stonewall Jackson), and "Better Foods, Better Homes Week," which featured daily musical concerts and lecture-demonstrations for women, sponsored by the *Charlotte Observer*.[25] The annual report for 1914 (published in the *Charlotte News* on January 23, 1915) lauded the fact that the Charlotte YMCA had in part fulfilled its educational mission by hosting seventy-two lectures in 1914, most of which were illustrated with stereopticon slides, though a number featured moving pictures, including titles identified as *Panama Canal, Inauguration of President Wilson*, and *Asphaltum Industry.* The three NAM films shown to the Telephone Society in March 1915 had, as I noted, also been screened, under the auspices of the YMCA's "educational committee,"[26] at this same facility a year earlier. On at least one other occasion in 1915, motion pictures were screened at the YMCA—again connected with telephony—when the Southern Bell Telephone & Telegraph Company sponsored a free afternoon screening of its titles, *The Girl at Central* and *The Operation of a Modern Telephone Exchange*.[27]

The process of re-searching digital newspaper archives that I have described in this abbreviated way could be extended further, by pursuing, for instance, *The Girl at Central* and *The Operation of a Modern Telephone Exchange,* the YMCA's role as an exhibition site and sponsor of screenings across the United States, or the film-related activities of the National Association of Manufacturers or Southern Bell Telephone & Telegraph. I will conclude, however, by briefly noting what the use of other digital archives that focus on print material apart from newspapers might bring to this project.

As would be expected, given its purview, searches of the Media History Digital Library yield little if any information about telephone societies or the YMCA in the mid-1910s. But this resource contains more material when it comes to the three NAM films. For example, Edison ads in the *Moving Picture World* from 1912 identify *The Crime of Carelessness* and *The Workman's Lesson* as hav-

ing been "produced in co-operation with the National Association of Manufacturers." An article from this trade magazine in December 1914 notes that *The Crime of Carelessness* has "been shown generally" in Indiana theaters in addition to the regular program "at the request" of the state fire marshal department. That same month, an exhibitor's column notes that the University of Kansas was also circulating all three NAM films, which were screened at the Opera House in Esbon, Kansas, "in connection with the regular program."[28]

A more relevant online resource, at least for this particular project, is the HathiTrust Digital Library, which has extensive holdings for books, periodicals, and pamphlets published before 1922 and therefore unambiguously in the public domain. Searches here lead, for example, both to Edison's initial promotional announcement for *The Crime of Carelessness* in the *Kinetogram*, his company's own "semi-monthly bulletin of Moving Picture News," and to the listing for the NAM films in the YMCA's 1916 handbook *Among Industrial Workers*, which explains that a range of titles are available for free to local YMCAs via the Motion Picture Bureau of the International Committee's Industrial Department.[29] Potentially even more interesting is the access through the HathiTrust to the vast amount of information (including advertisements) found in trade magazines like *Telephone Review* and *Telephony: The American Telephone Journal* as well as in house organs like *Illinois Bell Magazine* and *Western Electric News*. These periodicals often provide quite detailed accounts of the activities of telephone societies and how motion pictures and illustrated lectures figured as part of the advertising and public relations efforts of Bell Telephone. Particularly in terms of its holdings in trade journals, ephemeral pamphlets, and in-house publications, the HathiTrust Digital Library has potentially considerable import for the historical study of nontheatrical cinema (and media history more generally). Using this resource in addition to the digital newspaper archives would mean generating more instances to take into account, more

leads to pursue, more potential ways to branch out, all further expanding the possibilities for re-searching (and re-searching). Even the small-scale process I have presented here of working outward from the March 18, 1915, meeting of the Charlotte telephone society hints at some of the difficulties as well as the opportunities for using digital archival resources to build a history of nontheatrical film from the bottom (or instance) up. With more than a billion pages re-searchable, there's the danger of drowning in minutiae, of always being able to frame another search, of mistaking a thousand pinpointable instances for a map, of assuming that an enumeration of sites explains the historical specificity, variety, and significance of multi-sited cinema in the United States in 1915. That is the challenge for me of having access to searchable primary print material and attending to the particularities of small data. At the same time, I am convinced that for historians of cinema in the broad, inclusive sense—especially over its first half-century—there is no going back to some other side of what increasingly looks to be a digital divide.

ENDNOTES

1 Eric Hoyt, "Lenses for Lantern: Data Mining, Visualization, and Excavating Film History's Neglected Sources," *Film History* 26, no.2 (2014), 146–68.

2 André Gaudreault, *Film and Attraction: From Kinematography to Cinema [2008]*, trans. Timothy Bernard (Urbana: University of Illinois Press, 2011).

3 *Main Street Amusements: Movies and Commercial Entertainment in a Southern City, 1895–1930* (Washington, DC: Smithsonian Institution Press, 1995).

4 Genealogybank.com is a subsidiary of NewsBank Inc., which provides online resources to libraries; newspaperarchive.com is owned by Heritage Microfilm; newspapers.com was launched in November 2012 by Ancestry.com, then billing itself as the "the world's largest online family history resource" based in Provo, Utah, with "localized" websites and corporate offices in eight countries including the United States.

5 "Journalists to Use 'Movies'," *Evening Missourian* [Columbia, Missouri], March 18, 1915: 3.

6 "Envoy to Present Movies," *Washington Post*, March 18, 1915: 3.

7 "Film Lecture Given on Norway, Sweden, Denmark," *Racine [Wisconsin] Journal-News*, March 18, 1915: 9; "Movies Teaching Harvard Players Basketball Tricks," *Bridgeport [Connecticut] Evening Farmer* March 18, 1915: 7; "St. Patrick Keeps Sweitzer on Jump," *Chicago Daily Tribune*, March 18, 1915: 13; "To Determine Why People Go Insane," *Poughkeepsie [New York] Eagle-News*, March 18, 1915: 7; "Good Talk on Minnesota," *Williamsburg [Iowa] Journal* Tribune, March 18, 1915: 4.

8 For a sense of the history of theatrical venues screening motion pictures in Charlotte, North Carolina, see Robert Allen's ambitious digital resource, *Going to the Show: Mapping Moviegoing in North Carolina,* http://docsouth.unc.edu/gtts/.

9 "The Telephone Society Meets Tonight at 8 o'Clock," *Charlotte [North Carolina] News*, March 16, 1915: 2; "Telephone Society of Charlotte Had Delightful Meeting Wednesday Night," *Charlotte [North Carolina] News*, March 18, 1915: 3.

10 I consider these films and the extensive motion picture activities of the National Association of Manufacturers more generally in "Locating Early Non-Theatrical Audiences," in *Audiences: Defining and Researching Screen Entertainment Reception*, ed. Ian Christie (Amsterdam: Amsterdam University Press, 2012), 81–95, 248–53.

11 *The Man He Might Have Been* is mentioned as playing, for example, at theaters in Abilene, Kansas (*Abilene Daily Chronicle*, March 12, 1913: 1); Phoenix, Arizona (*Arizona Republican*, March 14, 1913: 7); Chillicothe, Missouri (*Chillicothe Daily Constitution*, March 31, 1913: 4); Great Bend, Kansas (*Great Bend Tribune*, April 28, 1913: 1); Bryan, Texas (*Bryan Daily Eagle*, May 17, 1913: 5); and Seattle, Washington (*Seattle Star*, May 24, 1913: 6).

12 *East Oregonian* [Pendleton, Oregon], February 22, 1913: 5; *East Oregonian*, January 30, 1913: 5; *East Oregonian,* July 27, 1912: 5.

13 "To Show Latest Method of Accident Prevention," *Poughkeepsie [New York] Eagle-News*, July 14, 1915: 6; "New Citizens Told of Duty to Nation," *Evening Star [Washington DC*], July 13, 1916: 3; *Topeka [Kansas]*

State Journal, October 21, 1912: 5.

14 *Nevada State Journal* [Reno, Nevada], June 23, 1914: 6; *Rock Island [Illinois] Argus*, February 7, 1916: 2.

15 *Bridgeport [Connecticut] Evening Farmer*, December 18, 1913: 2

16 "'Movie' Circuit Idea Explained," *Daily Gate City* [Keokuk, Iowa], March 26, 1915: 8; "Brings Several Boxes of Films," *Daily Gate City* [Keokuk, Iowa], March 24, 1915: 7; *Evening-Times Republican* [Marshalltown, Iowa] March 27, 1915: 8.

17 *Charlotte [North Carolina] News*, February 18, 1914: 4; "Annual Report of Sec. Probert of Y.M.C.A.," *Charlotte [North Carolina] News*, January 23, 1915: 2.

18 "Industrial Betterment," *Wilkes-Barre [Pennsylvania] Record*, January 17, 1914: 6; "Moving Pictures for the Mills," *Anderson [South Carolina] Daily Intelligencer*, March 27, 1914: 7.

19 "Telephone Society Organized," *Charlotte [North Carolina] News*, November 4, 1914: 5.

20 "First Meeting of the Telephone Society Tonight," *Charlotte [North Carolina] News*, November 17, 1914: 2.

21 Charlotte Telephone Society Tomorrow 8 P.M." *Charlotte [North Carolina] News*, December 14, 1914: 2; "Telephone Society," *Charlotte [North Carolina] Observer*, December 15, 1914: 5; "E. L. Travis Speaks to Telephone Society," *Charlotte [North Carolina] News*, May 26, 1915: 12.

22 "Telephone Men Have Banquet," *Charlotte [North Carolina] News*, December 22, 1915: 3

23 "Telephone Folks Have Play Tonight," *Harrisburg [Pennsylvania] Telegraph*, September 21, 1915: 12; "Employees See 'Movies' of Telephone Workers," *Washington [DC] Times*, October 29, 1915: 12; "Films Show Running of Coast-to-Coast Line," *Washington[DC] Herald*, October 29, 1915: 3.

24 See, for example, "Troop N. 2 Boy Scouts Will Meet Saturday," *Charlotte [North Carolina] News* May 21, 1915: 13; "Ministerial Association to Meet Mondays," *Charlotte [North Carolina] News* April 3, 1915: 11.

25 On Lee-Jackson Day, *Charlotte [North Carolina] Observer*, October 10, 1915: 13; ad for musical recital, *Charlotte [North Carolina] News*, March 7, 1915: 15; on "Better Foods, Better Homes Week, "*Charlotte [North Carolina] Observer*, October 10, 1915: 13.

26 "Annual Report of Sec. Probert of Y.M.C.A.," *Charlotte [North Carolina] News*, January 23, 1915: 2; description of screening of NAM films, *Charlotte [North Carolina] News*, February 18, 1914: 4.

27 "Telephone Operation to be Shown in 'Movies' at Hanna Hall," *Charlotte [North Carolina] News*, March 28, 1915: 6.

28 Edison ads, *Moving Picture World* 12.12 (June 22, 1912): 1100; and *Moving Picture World* 14.12 (December 21, 1912): 1159; "Fire Prevention Discussed," *Moving Picture World* 22.12 (December 19, 1914): 1699; *Moving Picture World* 22.12 (December 12, 1914): 1550.

29 *Kinetogram* 7.10 (December 15, 1912): 15; *Among Industrial Workers (Ways and Means): A Handbook for Associations in Industrial Fields* (New York: International Committee of the Young Men's Christian Associations, 1916), 76.

BIBLIOGRAPHY

"Fire Prevention Discussed." *Moving Picture World* 22, no. 12 (December 19, 1914): 1699.

Gaudreault, André. *Film and Attraction: From Kinematography to Cinema.* Translated by Timothy Bernard. Urbana: University of Illinois Press, 2011.

Going to the Show: Mapping Moviegoing in North Carolina. http://docsouth.unc.edu/gtts/.

Hoyt, Eric. "Lenses for Lantern: Data Mining, Visualization, and Excavating Film History's Neglected Sources." *Film History* 26, no. 2 (2014): 146–68.

Kinetogram 7, no. 10 (December 15, 1912).

Moving Picture World 12, no. 12 (June 22, 1912).

Moving Picture World 14, no. 12 (December 21, 1912).

Waller, Gregory. "Locating Early Non-Theatrical Audiences." *Audiences: Defining and Researching Screen Entertainment Reception*, edited by Ian Christie, 81–95. Amsterdam: Amsterdam University Press, 2012.

Waller, Gregory. *Main Street Amusements: Movies and Commercial Entertainment in a Southern City, 1895–1930.* Washington, DC: Smithsonian Institution Press, 1995.

Young Men's Christian Associations of North America. *Among Industrial*

Workers (Ways and Means): A Handbook for Associations in Industrial Fields. New York: International Committee of the Young Men's Christian Associations, 1916.

USING DIGITAL MAPS TO INVESTIGATE CINEMA HISTORY

Laura Horak

In the basement of the Swedish Film Institute, a colleague pulled out a dusty, yellowing book filled with film titles, dates, numbers, and city names—a ledger of all the places Swedish films were sent during the First World War, how many prints were sent, and when.[1] However, in trying to untangle how exactly these films circulated, I found that I had discovered too much—too many dates, too many places, and too many other variables. How was I, or any researcher, to make sense of it all? The ungainliness of this mass of information led me to explore new technologies of mapping and spatial analysis.[2]

Digital geospatial technologies allow scholars to ask new kinds of questions, forge new kinds of collaborations, and present archival materials and discoveries in new ways. However, digital geospatial technologies also introduce new conceptual and practical challenges, from questioning these systems' epistemological frameworks to figuring out exactly what kind of information to display and how. This essay offers an overview of what film scholars have done with digital mapping tools, other spatial humanities projects we could learn from, and ideas for things to try in the future. It also provides tips about which tools to use for which tasks, and sketches conceptual and practical concerns to keep in mind when undertaking a digital mapping project.

ORIGINS

Geospatial technologies were designed for purposes quite different from what scholars in the humanities are now doing with them. Geographic Information Systems (GIS), powerful software packages for modeling and analyzing the attributes of a particular piece of land, were first developed by and for land use managers and earth scientists. GIS software allows users to create math-

ematical models of an area's topography, waterways, flora, fauna, built environment, and human population and layer them on top of each other in order to determine, for example, where to build a new settlement or which areas are at most risk from flooding. GIS first appeared in the early 1960s, introduced by the Canadian Federal Department of Forestry and Rural Development and the Harvard Laboratory of Computer Graphics.[3] The company Esri (Environmental Systems Research Institute, Inc.) put the first commercial GIS on the market in 1982, which was re-engineered and renamed ArcGIS in the late 1990s.[4]

Other digital mapping technologies became ubiquitous in the 1990s. The first web-based map viewer, PARC Map Viewer, was launched in 1993, followed by MapQuest in 1996, and Google Maps and Google Earth in 2005.[5] Though the US Department of Defense first developed satellite-based global positioning systems (GPS) in 1960, they did not permit civilians to use precision GPS until 2000.[6] GPS navigators like Garmin became widely available in the mid-2000s, as did GPS-enabled cell phones shortly thereafter.[7] Today, digital geospatial technologies affect many aspects of our lives, in ways both seen and unseen. (By 2013, there were around a billion searches on Google Maps every day.)[8] Some humanities scholars have been quick to explore how these technologies could allow new kinds of questions to be asked in their fields.

DIGITAL MAPPING IN THE HUMANITIES

Humanities scholars began investigating the possibilities of geospatial technology in the 1990s. In 1997, Canadian geographer Fraser Taylor called for "cybercartography" that would bring together people from different disciplines to create networked, multimedia, interactive, and socially focused ways of modeling and analyzing space. He declared that new digital technologies should not be limited to "the hands of governments, the military [or] big business."[9] In 2003 Taylor began to implement his vision with a team of researchers at Carleton University, creating the

Cybercartographic Atlas of Antarctica and the Cybercartographic Atlas of Canada's Trade with the World.[10] In 2007, one of Taylor's collaborators, Sebastién Caquard, began producing cinematic cybercartographies.

Some historians began using GIS in the mid-1990s, a practice that came to be called historical GIS.[11] Historians mainly used GIS to investigate histories of land use and transportation networks, reconstruct past landscapes, and create datasets that government analysts and social scientists could use.[12] Around the same time, Franco Moretti called for a "geography of literature," insisting that "geography is not an inert container . . . but an active force, that pervades the literary field and shapes it in depth."[13] He began experimenting with mapping "space in literature" (e.g., Balzac's version of Paris) and "literature in space" (e.g., the European diffusion of Don Quixote).[14] Some film historians also began experimenting with GIS during this period.

Geographers, historians, literature scholars, and others played with geospatial technologies in the first decades of the 2000s, but they remained somewhat isolated from each other. (For example, a geographer complained that Moretti's otherwise interesting work ignored existing work on literary geography published by geographers.)[15] At this same time, the field long called "humanities computing" was renamed "digital humanities" and it began to grow in new directions. In recent years, a subfield of digital humanities has developed called spatial humanities. The term was popularized by David J. Bodenhamer, John Corrigan, and Trevor M. Harris's 2010 anthology *The Spatial Humanities: GIS and the Future of Humanities Scholarship*[16] and the Spatial Humanities website created by the Scholars' Lab at the University of Virginia. Spatial humanities broadens earlier humanities-based GIS work by drawing from a wider range of disciplines, using GIS alongside other spatial technologies like GPS-enabled mobile phones, virtual reality, and gaming, and asking how humanities-based ways of

knowing can change the way geospatial technologies work. Spatial humanities teams have begun developing mapping software designed specifically for humanities scholars, teachers, and students (some of which I list in a table at the end of this article). It has never been easier to use digital mapping and spatial analysis tools to investigate the intertwined people, places, and events of the past, and the places constructed by creative works.

FILM STUDIES GEOSPATIAL PROJECTS

Films scholars have used digital geospatial technologies to create enriched maps of particular places to better understand what it was like to live there and go to the movies and to investigate film exhibition spaces (where they were, what they looked like, who owned them, who built them, and what factors made them succeed or fail). In addition, scholars have mapped the locations represented in films and the movements of films and exhibitors through space, and they have compared the reception of a particular film in different places. In "Mapping the Movies" (2009), Deb Verhoeven, Kate Bowles, and Colin Arrowsmith argue that GIS enables us to investigate movies as "one element in a place-based cultural performance" and "impel[s] us to decentre the cinema, revealing it as one place amidst other locations, one moment in the busy context of everyday life."[17] In 2009, Sébastian Caquard and Fraser Taylor also called for "cinematic cartography," an investigation of the ways that cinema works cartographically, of how cartography could be considered cinematically, and of other new ways of bringing movies and maps together.[18] Caquard and his collaborators pointed out things cartographers could learn from the history and theory of film, particularly reflexivity, sound/image interactions, and motion.[19]

Many film history mapping projects have fallen under the rubric New Cinema History, a term coined by Verhoeven, Bowles, Richard Maltby, and Mike Walsh in 2011 to describe a methodological focus on movie-going as a social and economic practice embedded

within spaces, communities, and material networks.[20] The international History of Moviegoing, Exhibition and Reception (HoMER) Network has been at the forefront of this research. Their website is a clearinghouse of film-related geospatial mapping projects and datasets.[21] Additionally, two new anthologies, *Explorations in New Cinema History* (2011) and *Locating the Moving Image* (2014), provide many excellent examples of film historical scholarship that takes advantage of GIS technology.[22]

The first film history mapping projects created enriched maps of cinema venues in particular places. In 2003, Jeffrey Klenotic began building a GIS model of venues in New Hampshire called Mapping Movies.[23] The tool allows you to locate venues on a map overlaid with information about topography, territorial boundaries, roads, railways, and population density. In a recent article, Klenotic uses this tool, alongside more traditional forms of archival evidence, to explore "the social, cultural, economic, and physical terrain" that enabled a woman to become the lone film exhibitor in her small New Hampshire town in the mid-1910s.[24]

Since 2008, Robert C. Allen has been building a GIS model of film exhibition in North Carolina called Going to the Show. Allen uses detailed Sanborn fire insurance maps as a base upon which icons representing cinema venues are positioned.[25] Clicking an icon reveals the venue's name, address, dates of operation, proprietors, capacity, "racial policy," and other information, as well as links to architectural drawings, newspaper clippings, and postcards. Allen argues that Going to the Show reveals "a social geography of early moviegoing in North Carolina." His GIS model suggests that these theaters were "tightly woven" into the towns' civic life, not operating as an alternative, working-class sphere.[26]

Some projects have much larger scopes, such as Cinema Treasures, a crowd-sourced database of historical and contemporary cinemas that can be plotted on a Google map by country, and the

Cinematographic Atlas of Canadian Movie Theatres, which similarly plots Canadian cinemas onto a Google map. Where the city and state-based projects focus on the detailed physical, economic, and demographic contours of a particular area, larger maps like these are quite abstract. Pins or dots representing movie theaters float in the otherwise empty green and beige expanse of the standard Google base map, broken up only by major political boundaries. Where and how the pins cluster reveals the national-scale distribution of cinema theatres, which, not surprisingly, tends to follow population density. One virtue of these maps is that you can easily shift from a very large perspective to a very detailed one by zooming until you can see which particular blocks the cinemas are located on. However, these detailed views have nowhere near the detail of the enriched maps described above. It is not yet clear how one could generate a richly textured global or national map of cinema theaters that does not verge on incomprehensibility or how best to create a representation that works at both global and local scales and all the perspectives in between.

New projects open out beyond cinema to broader cultural life. One example is film scholar Annie Fee's new project, A Counter-Cartography of Paris Film Culture, 1918–1925, which will produce a "deep map" of the experiences of non-elite film viewers in Paris after the First World War.[27] The map will integrate cinema locations, autobiographical accounts, news stories, and details of the political meetings and protests held within these cinemas. Users can manipulate the variables to see, for example, the working-class cinemas a popular film series played in, as well as the cinemas used to organize industrial strikes in the same period, in order to visualize "the spatio-cultural boundaries of Parisian cinema practices during a time of great political and cultural upheaval."

"Deep mapping" holds particular promise for film historians. Initially suggested by the Situationist International in 1950s France and popularized by William Least Heat-Moon in the 1990s, deep

mapping "attempts to record and represent the grain and patina of a place through juxtapositions and interpenetrations of the historical and the contemporary, the political and the poetic, the discursive and the sensual."[28] Historian David J. Bodenhamer writes that "deep mapping conflates oral testimony, anthology, memoir, biography, images, natural history and everything you might ever want to say about a place."[29] He continues: "The deep map is meant to be visual and experiential, immersing users in a virtual world in which uncertainty, ambiguity, and contingency are ever-present, influenced by what was known (or believed) about the past and what was hoped for or feared in the future."[30] Bodenhamer argues that layers of a deep map should be "opened, wiki-like, to anyone with a memory or artefact to contribute," so that they become "a conversation and not a statement," a contested and contestable space of communal memory.[31]

In addition to mapping film venues within cultural and material landscapes, film scholars have also begun mapping cinematic representations of particular places. Mapping the City in Film, set in Liverpool, plots amateur, documentary, and newsreel films shot in Liverpool, as well as film venues, onto a map of the city. The project has also created a database of films shot in Liverpool that can by searched by type of "spatial representation" (e.g., public buildings and spaces, housing, maritime, etc.) or "spatial usage" (e.g., civil, commercial and industrial, festivals and parades, etc.). Likewise, Denmark on Film and Britain on Film allow you to click on pins stuck into maps to see clips from nonfiction films shot in those locations. In the Danish project, users can also upload their own films and add information about the films already posted, potentially turning the project into the kind of deep map Bodenhamer calls for.

While these projects map short, nonfiction films, Caquard's Cybercartographic Atlas of Canadian Cinema maps feature-length fiction films, which requires a more complex set of translations. In

conversation with mapping projects in literary studies, Caquard and his collaborator Jean-Pierre Fiset write that narrative mapping allows scholars "to explore the geographic structure of a story, and to better understand the impact of stories on the production of places."[32] So far, Caquard and his collaborators have made animated maps of forty-six Canadian feature films. Using points and lines, the atlas identifies the places where action unfolds, places mentioned by characters, and connections between the places.[33] Animation conveys the temporal dimension of the unfolding story. Caquard and Daniel Naud have described four key cartographic shapes formed by these narratives and hope to add more types of information to their model.[34]

Most of these projects take the cinematic representation of place quite straightforwardly. They do not address place substitution or the varied strategies of constructing space and place in film (via sets, special effects, editing, etc.).[35] It would be interesting to try to map practices of place substitution—for example, to visualize all the cities that Vancouver, BC, has stood in for, or all the places used to stand in for New York City.[36] If we added a temporal component to this map, we could see the effects of tax rebates and other industrial and political forces. Might there also be a way to map the complex construction of place in a film that, for example, is set in city A, shoots some exteriors in city B, shoots other exteriors and all interiors in city C, and then hires special effects companies in cities D, E, and F to composite everything? What would this map look like? What could it tell us about the geographical flows of cultural production? María Velez-Serna, one of the researchers developing a digital map of early cinema in Scotland, describes the difficulties of mapping both shooting locations and diegetic film settings, as well as the "imprecise geographies" of creative works in their GIS project.[37] While many fan sites plot out the movie scenes shot in a particular city (e.g., MapHook's San Francisco Movie Map), one of the few projects that allows users to see both where movies were shot and where they were ostensibly set

is the Cultural Atlas of Australia. This kind of map suggests what parts of the country are considered cinematic, able to be narrated, and/or interesting and what parts are not. It would be interesting to add the dimension of time and see how these geographies have changed. What about fantasy locations that are shot in real places? Considering the varied cinematic strategies of constructing place could make our digital cinematic cartographies even richer.

TRACKING MOVEMENT

Digital geospatial technologies also have great potential for tracking the many flows that make up the cinema—the flow of raw materials (chemicals, film stock, ideas), film prints, digital cinema packages, publicity material, and people. Film has always been a traveling medium, yet often our approach to it has been confined to the boundaries of particular nations (or regions or cities). While some scholars have outlined aspects of distribution via written narratives and tables[38] and investigated how particular stars were received in different places,[39] trying to follow the complex movements of people and things has proved extremely difficult. Oftentimes comparative projects place accounts from different locations side-by-side rather than forging connections between them. To really trace these movements and connect global and local scales requires active transnational collaborations, funding, and new tools.[40] These kinds of investigations are important because we know that cinema is a crucial carrier for ideas about the world and that it mediates encounters between disparate people and places, but we still have only piecemeal ideas about which people encountered which films when and how they made sense of the films they encountered. Who got to watch what? Who didn't? What did it mean to get connected into particular circuits? How did these connections change local entertainment cultures?

Deb Verhoeven and Colin Arrowsmith were likely the first to use digital tools to better understand film distribution. In 2009 they

began mapping the midcentury Greek cinema circuit in Australia. Using animated maps, they discovered that Greek cinemas appeared in particular neighborhoods before an influx of Greek immigrants settled there and that the cinemas disappeared shortly before the immigrants dispersed.[41] However, when they started to trace the movements of particular films, they decided not to use maps. They instead used a statistical tool called a Markov chain, which they visualized using forking branches of an olive tree.[42]

More recently, Cinema Context, which contains extensive information about screenings, venues, production companies, and censorship in the Netherlands between 1896 and 1960, has begun allowing users to see all the places that a particular film has played in the Netherlands as points on a Google map. Similarly, film historian Paul S. Moore uses Google Maps to visualize the travels of early itinerant exhibitors in Canada and the United States. While these projects do track movement, they primarily consist of points on a map, rather than tracing connections between the points.

Flow maps offer a powerful approach to investigating movement that has not yet been taken up by film scholars. French cartographer and statistician Charles Joseph Minard pioneered flow maps in the mid-1800s. His most famous was of Napoleon's Russian Campaign of 1812, which managed to plot six separate variables on a single map (see figure 1). No less an expert in "graphical method" than Étienne-Jules Marey praised the map's "brutal eloquence," which "seems to defy the pen of the historian."[43] More recently, data visualization expert Edward Tufte declared that Minard's map "may well be the best statistical graphic ever drawn."[44] Can we learn from Minard's "brutal eloquence" how to map the complex movements of people and things in cinema history?

Film studies can also learn much from flow mapping projects taking place elsewhere in the spatial humanities and social sciences. The Center for Spatial Studies at the University of Redlands' web-

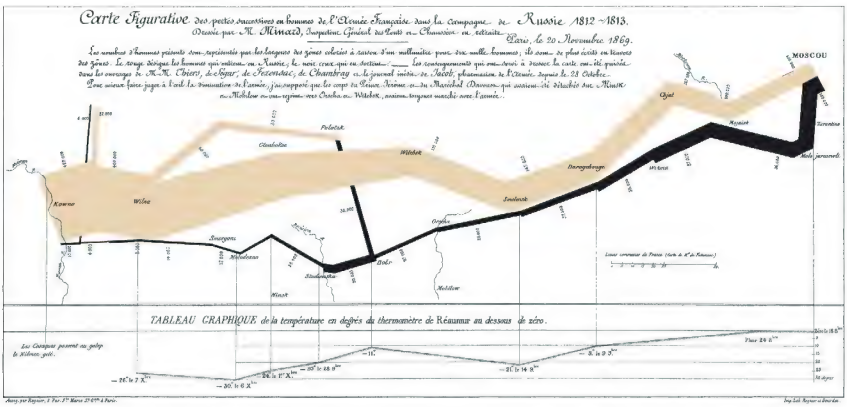


Figure 1. Charles Minard's 1869 chart showing the number of men in Napoleon's 1812 Russian campaign army, their movements, as well as the temperature they encountered on the return path, courtesy of Wikimedia Commons.

page on Flow and Movement offers a helpful introduction. There are at least four kinds of flow maps: point-to-point, distributive, network, and fields.[45] Point-to-point maps (also called origin-destination maps) show movement from point A to point B. When they center on a single origin (or destination), they are called "radial" maps. The Humanities+Design Lab at Stanford University's Mapping the Republic of Letters generates point-to-point maps of letters sent and received by Enlightenment thinkers such as Voltaire (see fig. 2). Clicking on a line reveals the author, source, destination, recipient, and date of the letter, and sometimes a link to the letter's contents. In contrast, *Forbes* magazine's interactive map of internal migration in the United States is a compelling example of a radial map.

Like radial maps, distributive maps (also called branching maps) typically have a single origin, but their lines stay joined together until they get close to their destinations. Minard created many distributive maps to portray the import or export of particular commodities (see fig. 3). The thickness of the branched line corresponded to the amount of the commodity. (In the map shown

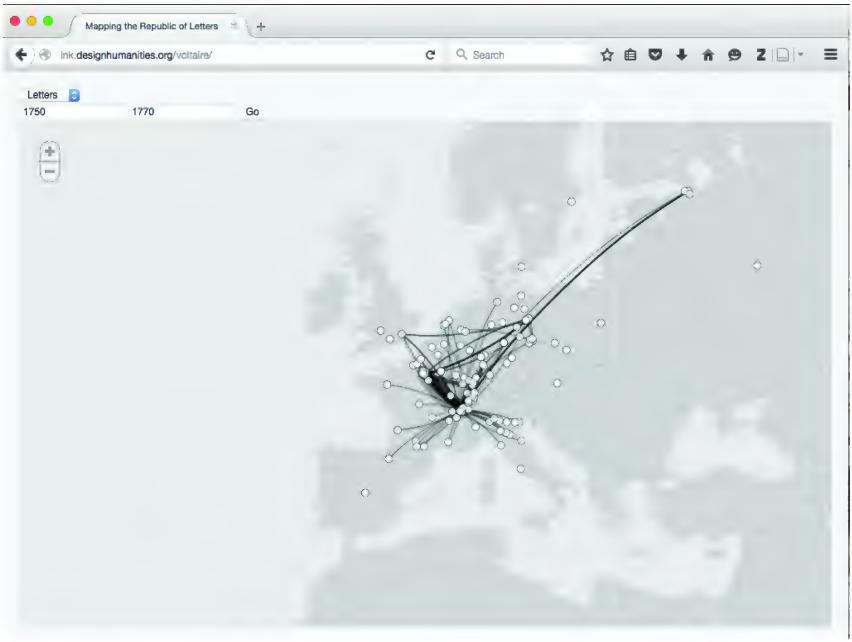


Figure 2. Representation of Voltaire's correspondence between 1750 and 1770 for Mapping the Republic of Letters.

here, each millimeter of thickness represents 20,000 tons of coal.) Though branching maps are not hard to make with ArcGIS and other software,[46] I have not found many examples of this type of map being used in the humanities.

Network maps don't have a single origin or destination but many interconnected points. They are often used to represent transportation and communication networks (e.g., roads, railways, telegraph cables, shipping routes, flight routes, etc.). One fascinating example is ORBIS: The Stanford Geospatial Network Model of the Roman World, an interactive GIS application that "reconstructs the duration and financial cost of travel in antiquity" (see fig. 4). In addition to exploring this immense network of roads, rivers, and sea routes, you can calculate the cheapest or fastest route from one place to another at a particular time, create a Minard-

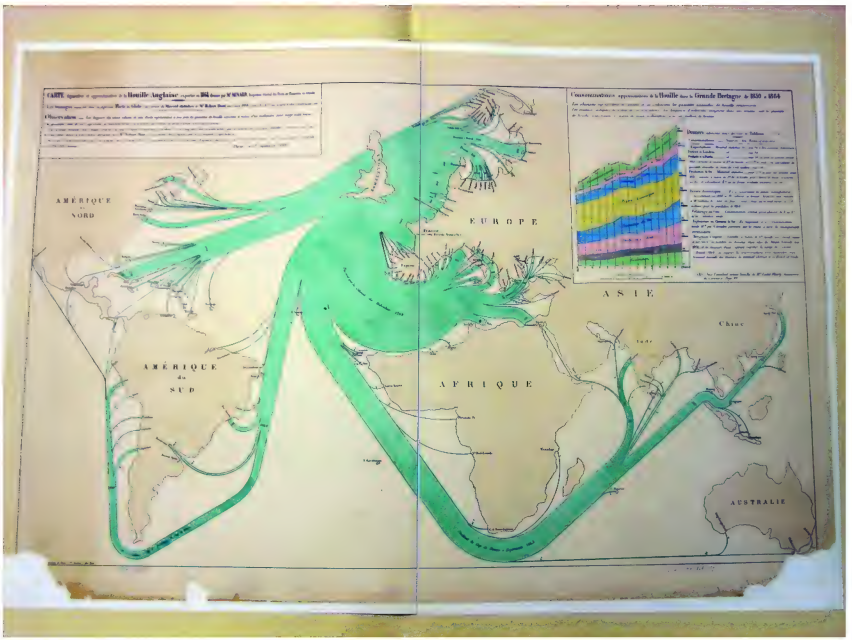


Figure 3. Charles Minard's 1890 map of British coal exports in 1864. From the Library of Congress, courtesy of Cartographia.

like distributive map showing the most used paths, and create a dynamic distance cartogram that shades areas of the map according to how expensive it is to get there from a particular point. This final option generates the fourth type of flow map, a field map. These maps represent a continuous surface of flow or movement not confined to particular channels. Perhaps the most dramatic example is computer programmer Cameron Beccario's real-time map of global wind and ocean currents (see fig. 5).

Other kinds of flow visualizations dispense with maps altogether. Taking the actual geography out of the representation allows you to represent more complex data, but you lose the sense of spatiality. One type of visualization is the circular plot, such as one used by the Global Flow of People project. Another possibility is the slopegraph, such as the one used by the PeopleMovin

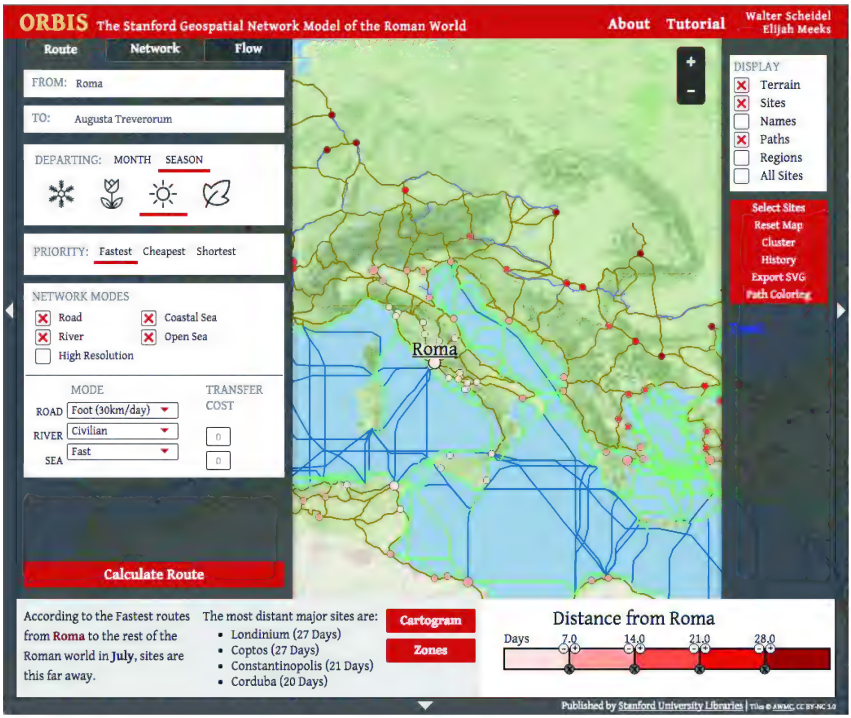


Figure 4. Land and sea networks of the Roman World on ORBIS.

project, which also visualizes global migration. All of these different possible flow maps could help us understand the movement of people and things in cinema history.

CHALLENGES IN DIGITAL MAPPING

There remain a number of challenges for digital geospatial technologies, namely, dealing with time, narrative, and sound. As many have pointed out, GIS is extremely good at dealing with space, but not very good at dealing with time. Different people have suggested different ways of incorporating time into GIS models.[47] One approach that should be of particular interest to cinema scholars is the creation of animated maps. As Caquard and Taylor point out, filmmakers were animating maps long before cartographers were.[48] Two powerful animated maps from the spatial hu-

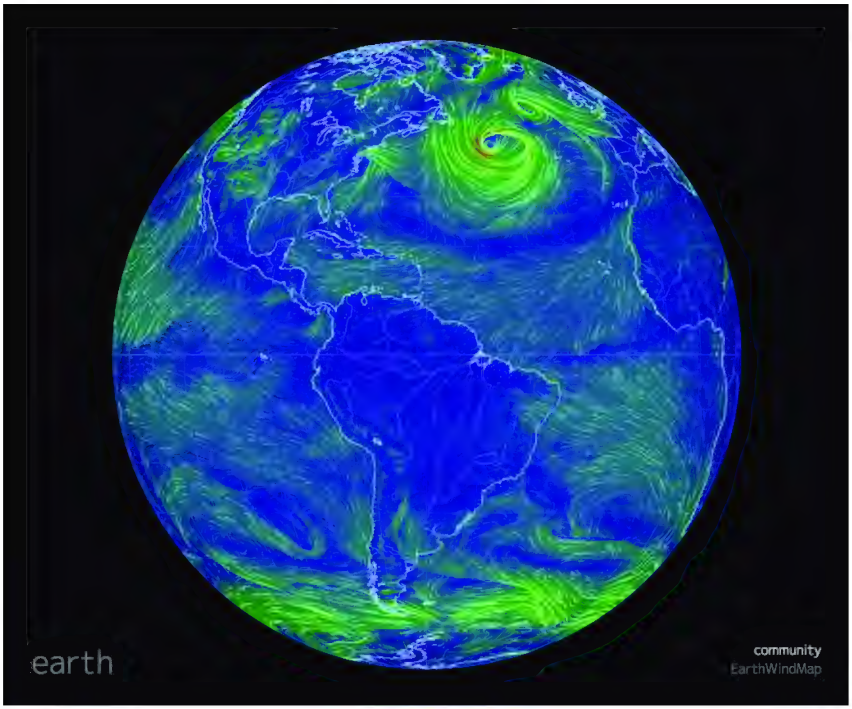


Figure 5. Field map of wind on earth's surface from EarthWindMap.

manities are Visualizing Emancipation by the Digital Scholarship Lab at the University of Richmond and Slave Revolt in Jamaica, 1760–1761: A Cartographic Narrative by Vincent Brown at the History Design Studio at Harvard University. These kinds of complex animated maps hold great promise for cinema history.

Related to the problem of time is the issue of narrative. Beyond attempts to map particular narratives, what is the relationship between mapping and narrative per se? Some mapping projects, like Slave Revolt in Jamaica, and mapping tools like Neatline and StoryMap aim to create new kinds of historical narratives—cartographic narratives.[49] Yet mapping could also offer an alternative to the strictures of narrative. In many projects, users can play around with layers, filters, and periods, making discoveries to the

side of the linear and causal structures that narratives inevitably reproduce. And yet, this freedom to play also contains the danger of incoherence and meaninglessness. There remain many open questions about how closely spatial humanities scholars want to reconcile new cartographic approaches with traditional historical narration.

With the exception of "sound maps" that plot recorded sound onto digital maps, most digital mapping projects don't use sound at all.[50] Incorporating this crucial dimension has the potential to enrich and even fundamentally change the user's experience of digital spatial models.[51]

Some scholars have raised more fundamental problems with digital mapping. Digital humanities scholar Johanna Drucker argues that, for humanities scholars, "graphical tools are a kind of intellectual Trojan horse, a vehicle through which assumptions about what constitutes information swarm with potent force."[52] While scholars in the humanities are trained to read texts and audiovisual media critically, Drucker writes, "So naturalized are the Google maps and bar charts generated from spread sheets that they pass as unquestioned representations of 'what is.'"[53] She argues that "the rendering of statistical information into graphical form gives it a simplicity and legibility that hides every aspect of the original interpretive framework on which the statistical data were constructed." Drucker calls upon scholars to "rethink digital tools for visualization on basic principles of the humanities." These principles are: "first, that the humanities are committed to the concept of knowledge as interpretation, and, second, that the apprehension of the phenomena of the physical, social, cultural world is through constructed and constitutive acts, not mechanistic or naturalistic realist representations of pre-existing or self-evident information."[54] Drucker encourages us to imagine new ways to visualize in the end product the processes of selection and interpretation that create our "data," to model space without the

tyranny of the Cartesian coordinate system, and to represent time without linearity. Her explanation of the many ways in which data are discursively constructed is an important riposte to those who argue that digital tools provide a new "empiricism" to overly interpretive disciplines.[55]

Likewise, cultural studies scholar Heather Zwicker points out: "As we know from even basic critiques of cartography, maps hide as much as they reveal. Maps are technologies of mastery that too often participate in the tyranny of the visible, organizing knowledge along unacknowledged co-ordinates—spatial co-ordinates, one might say, instead of place-based knowledge."[56] Zwicker asks: "How might we represent a meaning-rich place using digital means without assuming unfettered visibility, total mastery, and pure knowability? How can we draw a digital map that recognizes, rather than disavows, its own inadequacy?"[57] Her Edmonton Pipelines project explores possible responses to these questions. Scholars have also suggested ways of making digital maps more self-reflexive, sharing historical authority, developing maps as a form of "play," and ethically cocreating maps with disenfranchised communities.[58] Already in 2009, Verhoeven, Bowles, and Arrowsmith insightfully argued that "for mapping to be a productive development for film studies, it needs to work by engaging our imagination, and challenging our assumptions. . . . [Mapping] offers most when it raises new questions about spatial and temporal connectivity, rather than promising closure on the question of what was going on in the past."[59]

GETTING STARTED

I will conclude this article with some practical advice for starting to work with digital maps, based on the things that have helped me conceptualize my project on early Swedish film distribution. Edward Tufte's four volumes, from *The Quantitative Display of Information* (1983) to *Beautiful Evidence* (2006), demonstrate how to make visualizations of all kinds that are as clear and truthful as

possible (with many examples of visualizations that are not clear and not truthful).[60] Likewise, cartographer Mark Monmonier's *How to Lie with Maps* (1991) describes the codes and conventions of cartographic representation and the opportunities for distortion therein.[61] For an excellent introduction to the key concepts and terminology of GIS, see Karen K. Kemp's "Geographic Information Science and Spatial Analysis for the Humanities" (2010).[62] For a list of the kinds of questions to ask when designing a mapping project and the decisions made during the undertaking of one early project, the Orlando project, see the multiply authored "Mapping Tags and Tagging Maps" (2015).[63] At the end of the chapter you will find a table of some digital geospatial tools that have been recommended to me, divided into online platforms (including the most user-friendly options), programming libraries, and desktop software.

For my project mapping the global circulation of Swedish silent films, I am still collecting and organizing the data in a way that mapping software can interpret it. From what I have heard, this is often one of the most time consuming parts of the process. Once the data is ready, I will work with colleagues at Carleton's MacOdrum Library and Geomatics and Cybercartographic Research Center to do some trial visualizations to figure out which ways of mapping the data are most revealing. The process of figuring out what patterns we can find and how best to analyze and communicate them will be iterative. I want to experiment with representing the data as point-to-point, distributive, and network flow maps, and with different ways to represent change over time. I plan to try a few different tools, probably starting with ArcGIS, CartoDB, and Nunaliit, because they are powerful and flexible. Ultimately, I aspire to the "brutal eloquence" of the early years of print-based flow maps within a digital environment, through an iterative, playful, collaborative process.

Digital geospatial technologies will allow us to ask new questions

and revisit old assumptions about how space has shaped cinema history. They will also allow us to reach students and the public in new ways—not only to share our findings but also to share in the exhilaration of discovery, as the best spatial humanities projects do. However, even as we learn how to play with maps, we must also learn to be critical readers of maps and teach our students how to read these representations critically. At its best, mapping will not answer questions once and for all, but will reveal how many things we haven't even thought to ask yet.

Name and Website	What it's good at	Made by
ONLINE PLATFORMS & MAP DATA		
CartoDB, https://cartodb.com/	A cloud-based GIS platform. You can create maps, load data, customize visualizations, share, and collaborate with your colleagues. **Free + Subscription.**	CartoDB (Brooklyn, USA and Madrid, Spain)
Exhibit 2.0 (SIMIE Widget), http://www.simile-widgets.org/exhibit/	Publishing framework for data-rich interactive web pages. Exhibit lets you create web pages with with interactive maps, timelines, and other visualizations. **Free**.	David François Huynh (MIT), maintained by open source community
Google Earth + Google Earth Pro, http://www.google.ca/earth/explore/products/desktop.html	View satellite imagery, maps, terrain, 3D buildings, galaxies far in space, and the deepest depths of the ocean. Can create a narrated tour. **Free**.	Google (Mountain View, CA)
Google Maps API, http://www.google.com/get/mediatools/visualize.html#tab=api	Build customized, immersive maps. Access our map data then use Styled Maps to customize the display. You can also visualize your data with symbols and heatmaps. **Free**.	Google (Mountain View, CA)
Google My Maps, http://www.google.com/get/mediatools/visualize.html#tab=engine	Create a custom, interactive map in minutes with My Maps, which lets you draw and style points of interest, lines, and shapes on a map — no programming required. You can import map data from CSV, Excel, and KML files. **Free**.	Google (Mountain View, CA)

Google Tour Builder, https://tourbuilder.withgoogle.com/	Allows you to easily add photos and videos to a sequence of locations on Earth and then share links to these tours. **Free**.	Google (Mountain View, CA)
HyperCities, http://www.hypercities.com/	Thick mapping in the Digital Humanities. 3D reconstructions, multimedia, archival maps, social media feeds, and hypertext to create or explore historical narratives. Can be imported into a Scalar project. **Free**.	University of California, Los Angeles, University of Southern California, and City University of New York
MapScholar, http://mapscholar.org/	Enables humanities and social science scholars to create digital "atlases" featuring high-resolution images of historic maps. **Free**.	SHIVA, Scholars' Lab, IATH, Digital Media Lab, Digital Curation Services, and Department of History at the University of Virginia
Neatline, http://neatline.org/	Tell stories with maps and timelines. An add-on tool for Omeka, an open source web-publishing platform for the display of library, museum, archival, and scholarly collections and exhibitions. Recommends AcuGIS for hosting. **Free.**	The Scholars' Lab at the University of Virginia Library and the Roy Rosenzweig Center for History and New Media, George Mason University
OpenStreetMap, https://www.openstreetmap.org/	A free, editable map of the world that is being built by volunteers largely from scratch and released with an open-content license. **Free.**	Open source community and the OpenStreetMap Foundation (West Midlands, UK)
Palladio, http://palladio.designhumanities.org/#/	A web-based platform for visualizing complex, multidimensional data. In the Map view, you can see any coordinates data as points on a map. Relationships between points can be connected by lines, with the arc of the line representing the flow of the relationship. **Free**.	Humanities + Design Lab at Stanford

SimplyMap, http://www.simplymap.com/	Web-based mapping and data analysis application. Enables nontechnical and advanced users to quickly create professional thematic maps and reports with over 75,000 data variables.	Geographic Research, Inc. (New York, NY)
Story Maps by ArcGIS Online, http://storymaps.arcgis.com/en/	Application templates for ArcGIS that include: map-based tours, collections of points of interest, in-depth narratives, presenting multiple maps, and more.	Esri (Redlands, CA)
StoryMap JS, https://storymap.knightlab.com/	Allows you to tell stories on the web that highlight the locations of a series of events. Often used by online newspapers. **Free**.	Knight Lab at Northwestern University
VisualEyes + VisualEyes5, http://www.viseyes.org/	Weave together images, maps, charts, video, and data into interactive and compelling dynamic visualizations. **Free**.	SHANTI Interactive at University of Virginia
WorldMap, http://worldmap.harvard.edu/	Upload large datasets and overlay them with thousands of other layers, create and edit maps and link map features to rich media content, collaborate with small or large groups, make use of powerful online cartographic tools, georeference paper maps online, publish one's data to the world or to just a few collaborators. **Free**.	The Center for Geographic Analysis (CGA) at Harvard University
PROGRAMMING LANGUAGES/LIBRARIES		
D3.js, http://d3js.org/	(JavaScript) A library for manipulating documents based on data, using HTML, SVG, and CSS. Tools include DataMaps. **Free**.	Mike Bostock
ggplot2, http://ggplot2.org/	(R) An integrated suite of software facilities for data manipulation, calculation and graphical display, based on the grammar of graphics. **Free**.	The R Foundation

Kartograph, http://kartograph.org/	(Python and JavaScript) Simple and lightweight framework for building interactive map applications without Google Maps or any other mapping service. For designers and data journalists. **Free**.	Gregor Aisch (Graphics Editor of the New York Times)
Leaflet, http://leafletjs.com/	(Javascript) An open source library for mobile-friendly interactive maps. **Free**.	Vladimir Agafonkin of Mapbox
OpenLayers 3, http://openlayers.org/	(Javascript) An open source library to load, display, and render maps from multiple sources on web pages. **Free**.	MetaCarta
Processing, https://processing.org/	(Javascript and Python) A flexible software sketchbook and a language for learning how to code within the context of the visual arts. **Free**.	Processing Foundation, Fathom Information Design (Boston), the UCLA Arts Software Studio (Los Angeles), and ITP at NYU (New York City)
DESKTOP SOFTWARE		
ArcGIS, https://www.arcgis.com/	The grand-daddy of GIS software. Very powerful but expensive and difficult to learn. Tools include: Esri Story Map, WebGIS, Online Plan Routes, Distributive Flow Lines GP Tool, XY to Line Tool. Windows only.	Esri (Redlands, CA)
Flowmap, http://flowmap.geo.uu.nl/	A software package dedicated to analyzing and displaying interaction or flow data. Windows only. **Free**.	The Faculty of Geosciences of the Utrecht University in the Netherlands
GRASS GIS, https://grass.osgeo.org/	Geospatial data management and analysis, image processing, graphics and maps production, spatial modeling, and visualization. A general purpose raster/vector GIS combined with integrated image processing and data visualization subsystems. **Free**.	GRASS Development Team, including Construction Engineering Resesarch Laboratory (CERL) in Champaign, IL

Mapbox Studio Classic, https://www.mapbox.com/mapbox-studio-classic/#darwin	Design radically custom maps and datasets powered by vector tiles. (Formerly TileMill). **Free + Subscription**.	MapBox (an open-source company) based in Washington, DC, San Francisco, Ayacucho, Bangalore, and Berlin.
MapInfo Pro, http://www.pitneybowes.com/us/location-intelligence/geographic-information-systems/mapinfo-pro.html	A powerful mapping and geographic analysis application designed to show the relationship between data and geography in a more visual way.	Pitney Bowes (Stamford, CT)
MapWindow, http://www.mapwindow.org/	An open source desktop GIS with an extensible plugin architecture. **Free**.	Daniel P. Ames of Idaho State University Geospatial Software Lab and the MapWindow Open Source Team
Nunaliit Atlas Framework, http://nunaliit.org/	Create interactive mapping web sites based on your data and multimedia; allows web users to contribute additions and changes where permitted. **Free**.	The Geomatics and Cartographic Research Centre at Carleton University in Ottawa, Canada
QGIS, http://www.qgis.org/	An open source alternative to ArcGIS. Tools include: FlowMapper QGIS Plug-In v0.4, Oursins 1.0.1. **Free**.	Open source community

ENDNOTES

1 Thank you to Nadi Tofighian for bringing this ledger to my attention!

2 The ledger overturned historians' assumption that Swedish films first reached international screens during the "First Golden Age of Swedish Cinema" (1917–1924). Laura Horak, "The Global Distribution of Swedish Silent Film," in *The Blackwell Companion to Nordic Cinema*, ed. Mette Hjort and Ursula Lindqvist (West Sussex: Wiley-Blackwell, 2016), 457–84.

3 Pinde Fu and Jiulin Sun, *Web GIS: Principles and Applications*

(Redlands: ESRI Press, 2011), 4–5; David J. Bodenhamer, "The Potential of Spatial Humanities," in *The Spatial Humanities: GIS and the Future of Humanities Scholarship*, ed. David J. Bodenhamer, John Corrigan, and Trevor M. Harris (Bloomington: Indiana University Press, 2010), 16.

4 "About Esri: History Up Close," *Esri*, accessed November 27, 2015, http://www.esri.com/~/media/Files/Pdfs/about-esri/esri-history-up-close.

5 In 2004, Google acquired Keyhole, Where2, and Zipdash. Google Maps and Google Earth were based on the technologies these companies had developed. Fu and Sun, *Web GIS*, 7–8; Sarah Kessler, "MapQuest Founder on the Past, Present and Future of Location Tech," *Mashable*, November 30, 2010, http://mashable.com/2010/11/30/mapquest-founder-interview/; Liz Gannes, "Ten Years of Google Maps, From Slashdot to Ground Truth," *Re/code*, February 8, 2015, http://recode.net/2015/02/08/ten-years-of-google-maps-from-slashdot-to-ground-truth/.

6 Jamie Lendino, "The History of Car GPS Navigation," *PCMAG*, April 16, 2012, http://www.pcmag.com/article2/0,2817,2402755,00.asp.

7 Ibid.

8 Tom Chivers, "The Story of Google Maps," *Telegraph*, June 4, 2013, http://www.telegraph.co.uk/technology/google/10090014/The-story-of-Google-Maps.html.

9 D. R. Fraser Taylor, "Maps and Mapping in the Information Era," in *Proceedings of the 18th International Cartographic Conference*, ed. Lars Ottoson, vol. 1 (Gävle, Sweden, 1997), 1–10.

10 Sébastien Caquard and Fraser Taylor, "What Is Cinematic Cartography?," *Cartographic Journal* 46, no. 1 (2009): 5–8. These and other cybercartographic atlases can be viewed here: https://gcrc.carleton.ca/confluence/display/GCRCWEB/Atlases - Atlases-CTW.

11 Anne Kelly Knowles and Amy Hillier, "Preface," in *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship*, ed. Anne Kelly Knowles and Amy Hillier (Redlands: ESRI Press, 2008), xiii.

12 Anne Kelly Knowles, "GIS and History," in *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship*, ed. Anne Kelly Knowles (Redlands: ESRI Press, 2008), 1–25. For more information about Historical GIS, see Anne Kelly Knowles, ed., *Past Time, Past Place: GIS for History* (Redlands: ESRI Press, 2002); Anne Kelly Knowles, ed.,

Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship (Redlands: ESRI Press, 2008).

13 Franco Moretti, *Atlas of the European Novel, 1800–1900* (London: Verso, 1998), 3.

14 Ibid.

15 David Matless, "Book Review: An Atlas of the European Novel 1800–1900," *Progress in Human Geography* 23, no. 4 (1999): 659–60, doi:10.1191/030913299677607263.

16 David J. Bodenhamer, John Corrigan, and Trevor M. Harris, eds., *The Spatial Humanities: GIS and the Future of Humanities Scholarship* (Bloomington: Indiana University Press, 2010), 395. The book launched a "Spatial Humanities" series at Indiana University Press. Recent publications include: Ian N. Gregory and Alistair Geddes, eds., *Toward Spatial Humanities: Historical GIS and Spatial History* (Bloomington: Indiana University Press, 2014); David J. Bodenhamer, John Corrigan, and Trevor M. Harris, *Deep Maps and Spatial Narratives* (Bloomington: Indiana University Press, 2015).

17 Deb Verhoeven, Kate Bowles, and Colin Arrowsmith, "Mapping the Movies: Reflections on the Use of Geospatial Technologies for Historical Cinema Audience Research," in *Digital Tools in Media Studies: Analysis and Research: An Overview*, ed. Michael Ross, Manfred Grauer, and Bernd Freisleben (Bielefeld: Transcript, 2009), 70, 78.

18 Caquard and Taylor, "What Is Cinematic Cartography?" They exemplified these various approaches in a special edited volume of the *Cartographic Journal.* The term "cinematic cartography" inverts Tom Conley's concept of "cartographic cinema," which they cite as an inspiration.

19 Sébastien Caquard et al., "Designing Sound in Cybercartography: From Structured Cinematic Narratives to Unpredictable Sound/Image Interactions," *International Journal of Geographical Information Science* 22, no. 11–12 (2008): 1219–45, doi:10.1080/13658810801909649; Sébastien Caquard, "Foreshadowing Contemporary Digital Cartography: A Historical Review of Cinematic Maps in Films," *Cartographic Journal* 46, no. 1 (2009): 46–55, doi:10.1179/000870409X415589; Sébastien Caquard and Benjamin Wright, "Challenging the Digital Cartographic Continuity System: Lessons from Cinema," in *Cartography and Art*, ed. William Cart-

wright, Georg F. Gartner, and Antje Lehn, Lecture Notes in Geoinformation and Cartography (Berlin: Springer, 2009), 1–14.

20 Richard Maltby et al., *The New Cinema History: A Guide for Researchers* (Oxford: Wiley-Blackwell, 2011). *See also*: Richard Maltby, Daniël Biltereyst, and Philippe Meers, eds., *Explorations in New Cinema History: Approaches and Case Studies* (Malden: Wiley-Blackwell, 2011); Deb Verhoeven, "New Cinema History and the Computational Turn," *Beyond Art, Beyond Humanities, Beyond Technology: A New Creativity. Proceedings of the World Congress of Communication and the Arts Conference,* University of Minho, Portugal, 2012.

21 Many new mapping projects were presented at their recent conference "What Is Cinema History?"

22 Maltby, Biltereyst, and Meers, *Explorations in New Cinema History*; Julia Hallam and Les Roberts, eds., *Locating the Moving Image: New Approaches to Film and Place* (Bloomington: Indiana University Press, 2014).

23 In 2013 he migrated the system to a web-based GIS platform called Environmental Response Management Application (ERMA).

24 Jeffrey F. Klenotic, "Space, Place, and the Female Film Exhibitor: The Transformation of Cinema in Small-Town New Hampshire during the 1910s," in *Locating the Moving Image: New Approaches to Film and Place*, ed. Julia Hallam and Les Roberts (Bloomington: Indiana University Press, 2014), 45.

25 To fit a scanned archival map onto the coordinate system used by GIS models, you need to "georectify" the map by matching points on the map to points on a contemporary digital map. If the distances between points on the archival map are not the same as the distances between points on the digital map, the map will stretch and squish to fit (this is also called "rubber sheeting," because it makes the map look like a rubber sheet).

26 Robert C. Allen, "Reimagining the History of the Experience of Cinema in a Post-Moviegoing Age," in *Explorations in New Cinema History: Approaches and Case Studies*, ed. Richard Maltby, Daniël Biltereyst, and Philippe Meers (Malden: Wiley-Blackwell, 2011). *See also*: Robert C. Allen, "Getting to 'Going to the Show,'" in *Locating the Moving Image: New Approaches to Film and Place*, ed. Julia Hallam and Les Roberts (Bloom-

ington: Indiana University Press, 2014), 31–43.

27 Another good example is the "Mapping Irish Culture @ 1916: Dublin" project, by the Department of Media Studies at the National University of Ireland Maynooth.

28 William Least Heat-Moon, *PrairyErth: (a Deep Map)* (Boston: Houghton Mifflin, 1993), as quoted in Bodenhamer, "The Potential of Spatial Humanities," 26.

29 Bodenhamer, "The Potential of Spatial Humanities," 27. *See also* Bodenhamer, Corrigan, and Harris, *Deep Maps and Spatial Narratives*.

30 Bodenhamer, "The Potential of Spatial Humanities," 28.

31 Ibid., 26–28.

32 Sébastien Caquard and Jean-Pierre Fiset, "How Can We Map Stories? A Cybercartographic Application for Narrative Cartography," *Journal of Maps* 10, no. 1 (2014): 18, doi:10.1080/17445647.2013.847387.

33 For a detailed description, see: Sébastien Caquard and Daniel Naud, "A Spatial Typology of Cinematographic Narratives," in *Developments in the Theory and Practice of Cybercartography Applications and Indigenous Mapping*, ed. D. R. Fraser Taylor and Tracey P. Lauriault (Amsterdam and Boston: Elsevier Science, 2014), 164; Caquard and Fiset, "How Can We Map Stories?," 20.

34 Caquard and Naud, "A Spatial Typology of Cinematographic Narratives," 172.

35 Some excellent video essays visualize the cinematic construction of space without using maps, such as Timothy I. Smith's Mapping Cinematic Space: (Re)presentation of Hitchcock's *Rear Window* (2010) and Jeff Desom's *Rear Window* Timelapse (2012).

36 A great video essay on place substitution and Vancouver, BC, is Tony Zhou's Vancouver Never Plays Itself (2015). On place substitution more generally, see: Mark B. Sandberg, "Location, 'Location': On the Plausibility of Place Substitution," in *Silent Cinema and the Politics of Space*, ed. Jennifer M. Bean, Anupama Kapse, and Laura Horak (Bloomington: Indiana University Press, 2014), 23–46.

37 María Velez-Serna, "Mapping Settings and Locations," *Early Cinema in Scotland*, November 30, 2014, http://earlycinema.gla.ac.uk/mapping-settings-and-locations/. For more information on this project, see: http://

earlycinema.gla.ac.uk/.

38 Ivo Leopold Blom, *Jean Desmet and the Early Dutch Film Trade* (Amsterdam: Amsterdam University Press, 2003); Kristin Thompson, *Exporting Entertainment: America in the World Film Market, 1907–34* (London: BFI, 1985).

39 Marina Dahlquist, *Exporting Perilous Pauline: Pearl White and the Serial Film Craze* (Urbana: University of Illinois Press, 2013); Martin Loiperdinger and Uli Jung, eds., *Importing Asta Nielsen: The International Film Star in the Making, 1910–1914*, (Bloomington: John Libbey and Indiana University Press, 2013).

40 Part of tracing these connections involves bringing datasets collected in different places using different types of databases together. This is something that HoMER is actively working on.

41 Verhoeven, Bowles, and Arrowsmith, "Mapping the Movies: Reflections on the Use of Geospatial Technologies for Historical Cinema Audience Research," 76–77.

42 Deb Verhoeven and Colin Arrowsmith, "Mapping the Ill-Disciplined? Spatial Analyses and Historical Change in the Postwar Film Industry," in *Locating the Moving Image: New Approaches to Film and Place*, ed. Julia Hallam and Les Roberts (Bloomington: Indiana University Press, 2014), 106–29; Colin Arrowsmith and Deb Verhoeven, "Visual Methods for Showing Cinema Circuits at Varying Temporal and Spatial Scales" (GSR_1 Symposium, Melbourne, Australia, 2011).

43 Étienne-Jules Marey, *La Méthode graphique dans les sciences expérimentales* (Paris: G. Masson, 1885), 73. Translation mine. The original text reads: "Toujours il arrive à des effets saisissants, mais nulle part la representation graphique de la marche des armées n'atteint ce degré de brutale éloquence qui, dans la figure 37, semble défier la plume de l'historien."

44 Edward R. Tufte, *The Visual Display of Quantitative Information* (Cheshire: Graphics Press, 1983), 40. Cartographer Menno-Jan Kraak has asked even more recently if today's geospatial technologies can improve on Minard's map. Menno-Jan Kraak, *Mapping Time: Illustrated by Minard's Map of Napoleon's Russian Campaign of 1812* (Redlands: ESRI Press, 2014).

45 "Flow and Movement," *Center for Spatial Studies*, 2014, http://spatialstudies.redlands.edu/flow-and-movement/; Amanda Briney, "Overview of Flow Mapping," *GIS Lounge*, April 2, 2014, http://www.gislounge.com/overview-flow-mapping/.

46 For example, Kenneth Field has reproduced Minard's map of British Coal Exports in 1864 using ArcGIS, as seen here: http://arcg.is/1L0qOGJ.

47 Edward L. Ayers, "Mapping Time," in *GeoHumanities: Art, History, Text at the Edge of Place*, ed. Michael Dear et al. (London: Routledge, 2011), 215–25; Susan Brown et al., "Mapping Tags and Tagging Maps: Leveraging Spatial Markup for Literary History," in *Cultural Mapping and the Digital Sphere: Place and Space*, ed. Ruth Panofsky and Kathleen Kellett (Edmonton: University of Alberta Press, 2015), 3–24; Verhoeven, Bowles, and Arrowsmith, "Mapping the Movies: Reflections on the Use of Geospatial Technologies for Historical Cinema Audience Research"; Ian Gregory, "Exploiting Time and Space: A Challenge for GIS in the Digital Humanities," in *The Spatial Humanities: GIS and the Future of Humanities Scholarship*, ed. David J. Bodenhamer, John Corrigan, and Trevor M. Harris (Bloomington: Indiana University Press, 2010), 58–75; Kraak, *Mapping Time*; Simon McBride, Chris Bellman, and Colin Arrowsmith, "Developing Space-Time Models and Representation for GIS," in *Proceedings of the Spatial Sciences Coalition Conference* (Canberra, Australia, 2003).

48 Caquard and Taylor, "What Is Cinematic Cartography?"; Caquard, "Foreshadowing Contemporary Digital Cartography."

49 For more on narrative cartography and story maps, see: Sébastien Caquard and William Cartwright, "Narrative Cartography: From Mapping Stories to the Narrative of Maps and Mapping," *Cartographic Journal* 51, no. 2 (2014): 101–6, doi:10.1179/0008704114Z.000000000130; Sébastien Caquard, "Cartography I: Mapping Narrative Cartography," *Progress in Human Geography* 37, no. 1 (2013): 135–44, doi:10.1177/0309132511423796; Brian Greenspan et al., "Live Hypernarrative and Cybercartography: You Are Here, Now," *Cartographica: The International Journal for Geographic Information and Geovisualization* 41, no. 1 (2006): 35–46, doi:10.3138/L42H-7P81-NMX0-5540.

50 Paul Théberge, "Sound Maps: Music and Sound in Cybercartogra-

phy," in *Cybercartography — Theory and Practice*, ed. D. R. Fraser Taylor, vol. 4, Modern Cartography (Amsterdam: Elsevier Science, 2005), 389–410.

51 Caquard et al., "Designing Sound in Cybercartography."

52 Johanna Drucker, "Humanities Approaches to Graphical Display" 5, no. 1 (2011): para. 1, http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

53 Ibid.

54 Ibid., para. 7.

55 For example: Julia Hallam and Les Roberts, "Film and Spatiality: Outline of a New Empiricism," in *Locating the Moving Image: New Approaches to Film and Place*, ed. Julia Hallam and Les Roberts (Bloomington: Indiana University Press, 2014), 1–30.

56 Heather Zwicker, "Edmonton Pipelines: Living and Playing in the Digital City," in *Cultural Mapping and the Digital Sphere: Place and Space*, ed. Ruth Panofsky and Kathleen Kellett (Edmonton: University of Alberta Press, 2015), 134–35.

57 Ibid., 135.

58 Caquard and Wright, "Challenging the Digital Cartographic Continuity System"; Kate Bowles, "Beyond the Boundary: Vernacular Mapping and the Sharing of Historical Authority," in *Locating the Moving Image: New Approaches to Film and Place*, ed. Julia Hallam and Les Roberts (Bloomington: Indiana University Press, 2014), 221–44; Bethany Nowviskie, "How to Play with Maps," in *Cultural Mapping and the Digital Sphere: Place and Space*, ed. Ruth Panofsky and Kathleen Kellett (Edmonton: University of Alberta Press, 2015), 107–28; D. R. Fraser Taylor, "Some Recent Developments in the Theory and Practice of Cybercartography: Applications in Indigenous Mapping: An Introduction," in *Developments in the Theory and Practice of Cybercartography Applications and Indigenous Mapping*, ed. D. R. Fraser Taylor and Tracey P. Lauriault (Amsterdam: Elsevier Science, 2014), 2–15; Timothy di Leo Browne and Gita J. Ljubicic, "Considerations for Informed Consent in the Context of Online, Interactive, Atlas Creation," in *Developments in the Theory and Practice of Cybercartography: Applications and Indigenous Mapping*, ed. D. R. Fraser Taylor and Tracey P. Lauriault (Amsterdam: Elsevier Science, 2014), 263–77; Teresa Scassa, Tracey P. Lauriault, and D. R. Fraser Taylor,

"Cybercartography and Traditional Knowledge: Responding to Legal and Ethical Challenges," in *Developments in the Theory and Practice of Cybercartography: Applications and Indigenous Mapping*, ed. D. R. Fraser Taylor and Tracey P. Lauriault (Amsterdam: Elsevier Science, 2014), 279–97.

59 Verhoeven, Bowles, and Arrowsmith, "Mapping the Movies: Reflections on the Use of Geospatial Technologies for Historical Cinema Audience Research," 79.

60 Tufte, *The Visual Display of Quantitative Information*; Edward R. Tufte, *Envisioning Information* (Cheshire: Graphics Press, 1990); Edward R. Tufte, *Visual Explanations: Images and Quantities, Evidence and Narrative* (Cheshire: Graphics Press, 1997); Edward R. Tufte, *Beautiful Evidence* (Cheshire: Graphics Press, 2006).

61 Mark S. Monmonier, *How to Lie with Maps* (Chicago: University of Chicago Press, 1991). *See also*: Mark S. Monmonier, *Mapping It Out: Expository Cartography for the Humanities and Social Sciences* (Chicago: University of Chicago Press, 1993).

62 Karen K. Kemp, "Geographic Information Science and Spatial Analysis for the Humanities," in *The Spatial Humanities: GIS and the Future of Humanities Scholarship*, ed. David J. Bodenhamer, John Corrigan, and Trevor M. Harris (Bloomington: Indiana University Press, 2010), 31–57.

63 Brown et al., "Mapping Tags and Tagging Maps: Leveraging Spatial Markup for Literary History."

BIBLIOGRAPHY

"About Esri: History Up Close." *Esri*. Accessed November 27, 2015. http://www.esri.com/~/media/Files/Pdfs/about-esri/esri-history-up-close.

Allen, Robert C. "Getting to 'Going to the Show.'" In *Locating the Moving Image: New Approaches to Film and Place*, edited by Julia Hallam and Les Roberts, 31–43. Bloomington: Indiana University Press, 2014.

———. "Reimagining the History of the Experience of Cinema in a Post-Moviegoing Age." In *Explorations in New Cinema History: Approaches and Case Studies*, edited by Richard Maltby, Daniël Biltereyst, and Philippe Meers. Malden: Wiley-Blackwell, 2011.

Arrowsmith, Colin, and Deb Verhoeven. "Visual Methods for Showing

Cinema Circuits at Varying Temporal and Spatial Scales." GSR_1 Symposium Melbourne, Australia, 2011.

Ayers, Edward L. "Mapping Time." In *GeoHumanities: Art, History, Text at the Edge of Place*, edited by Michael Dear, Jim Ketchum, Sarah Luria, and Douglas Richardson, 215–25. London and New York: Routledge, 2011.

Blom, Ivo Leopold. *Jean Desmet and the Early Dutch Film Trade*. Amsterdam: Amsterdam University Press, 2003.

Bodenhamer, David J. "The Potential of Spatial Humanities." In *The Spatial Humanities: GIS and the Future of Humanities Scholarship*, edited by David J. Bodenhamer, John Corrigan, and Trevor M. Harris, 14–30. Bloomington: Indiana University Press, 2010.

Bodenhamer, David J, John Corrigan, and Trevor M. Harris. *Deep Maps and Spatial Narratives*. Bloomington: Indiana University Press, 2015.

Bodenhamer, David J., John Corrigan, and Trevor M. Harris, eds. *The Spatial Humanities: GIS and the Future of Humanities Scholarship*. Bloomington: Indiana University Press, 2010.

Bowles, Kate. "Beyond the Boundary: Vernacular Mapping and the Sharing of Historical Authority." In *Locating the Moving Image: New Approaches to Film and Place*, edited by Julia Hallam and Les Roberts, 221–44. Bloomington: Indiana University Press, 2014.

Briney, Amanda. "Overview of Flow Mapping." *GIS Lounge*, April 2, 2014. http://www.gislounge.com/overview-flow-mapping/.

Browne, Timothy di Leo, and Gita J. Ljubicic. "Considerations for Informed Consent in the Context of Online, Interactive, Atlas Creation." In *Developments in the Theory and Practice of Cybercartography: Applications and Indigenous Mapping*, edited by D. R. Fraser Taylor and Tracey P. Lauriault, 263–77. Amsterdam and Boston: Elsevier Science, 2014.

Brown, Susan, Isobel Grundy, Mariana Paredes-Olea, Jeffery Antoniuk, and Breanna Mroczek. "Mapping Tags and Tagging Maps: Leveraging Spatial Markup for Literary History." In *Cultural Mapping and the Digital Sphere: Place and Space*, edited by Ruth Panofsky and Kathleen Kellett, 3–24. Edmonton: University of Alberta Press, 2015.

Caquard, Sébastian, Glenn Brauen, Benjamin Wright, and Paul Jasen. "Designing Sound in Cybercartography: From Structured Cinematic Narratives to Unpredictable Sound/Image Interactions." *International Journal of Geographical Information Science* 22, no. 11–12 (November 1, 2008): 1219–45. doi:10.1080/13658810801909649.

Caquard, Sébastien. "Cartography I: Mapping Narrative Cartography." *Progress in Human Geography* 37, no. 1 (February 1, 2013): 135–44. doi:10.1177/0309132511423796.

———. "Foreshadowing Contemporary Digital Cartography: A Historical Review of Cinematic Maps in Films." *The Cartographic Journal* 46, no. 1 (February 1, 2009): 46–55. doi:10.1179/000870409X415589.

Caquard, Sébastien, and William Cartwright. "Narrative Cartography: From Mapping Stories to the Narrative of Maps and Mapping." *The Cartographic Journal* 51, no. 2 (May 1, 2014): 101–6. doi:10.1179/0008704114Z.000000000130.

Caquard, Sébastien, and Jean-Pierre Fiset. "How Can We Map Stories? A Cybercartographic Application for Narrative Cartography." *Journal of Maps* 10, no. 1 (January 2, 2014): 18–25. doi:10.1080/17445647.2013.847387.

Caquard, Sébastien, and Daniel Naud. "A Spatial Typology of Cinematographic Narratives." In *Developments in the Theory and Practice of Cybercartography Applications and Indigenous Mapping*, edited by D. R. Fraser Taylor and Tracey P. Lauriault, 161–73. Amsterdam and Boston: Elsevier Science, 2014.

Caquard, Sébastien, and Fraser Taylor. "What Is Cinematic Cartography?" *The Cartographic Journal* 46, no. 1 (2009): 5–8.

Caquard, Sébastien, and Benjamin Wright. "Challenging the Digital Cartographic Continuity System: Lessons from Cinema." In *Cartography and Art*, edited by William Cartwright, Georg F. Gartner, and Antje Lehn, 1–14. Lecture Notes in Geoinformation and Cartography. Berlin: Springer, 2009. http://link.springer.com/chapter/10.1007/978-3-540-68569-2_16.

Chivers, Tom. "The Story of Google Maps." *Telegraph*, June 4, 2013, Technology section. http://www.telegraph.co.uk/technology/google/10090014/The-storyof-Google-Maps.html.

Dahlquist, Marina. *Exporting Perilous Pauline: Pearl White and the Serial Film Craze*. Urbana: University of Illinois Press, 2013.

Drucker, Johanna. "Humanities Approaches to Graphical Display" 5, no. 1 (2011). http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

"Flow and Movement." *Center for Spatial Studies*, 2014. http://spatialstudies.redlands.edu/flow-and-movement/.

Fu, Pinde, and Jiulin Sun. *Web GIS: Principles and Applications*. Redlands: ESRI Press, 2011.

Gannes, Liz. "Ten Years of Google Maps, From Slashdot to Ground Truth." *Re/code*, February 8, 2015. http://recode.net/2015/02/08/ten-years-of-google-maps-from-slashdot-to-ground-truth/.

Greenspan, Brian, Claire Dormann, Sébastien Caquard, Chris Eaket, and Robert Biddle. "Live Hypernarrative and Cybercartography: You Are Here, Now." *Cartographica: The International Journal for Geographic Information and Geovisualization* 41, no. 1 (March 1, 2006): 35–46. doi:10.3138/L42H-7P81-NMX0-5540.

Gregory, Ian. "Exploiting Time and Space: A Challenge for GIS in the Digital Humanities." In *The Spatial Humanities: GIS and the Future of Humanities Scholarship*, edited by David J. Bodenhamer, John Corrigan, and Trevor M. Harris, 58–75. Bloomington: Indiana University Press, 2010.

Gregory, Ian N., and Alistair Geddes, eds. *Toward Spatial Humanities: Historical GIS and Spatial History*. Bloomington: Indiana University Press, 2014.

Hallam, Julia, and Les Roberts. "Film and Spatiality: Outline of a New Empiricism." In *Locating the Moving Image: New Approaches to Film and Place*, edited by Julia Hallam and Les Roberts, 1–30. Bloomington: Indiana University Press, 2014.

———, eds. *Locating the Moving Image: New Approaches to Film and Place*. Bloomington: Indiana University Press, 2014.

Heat-Moon, William Least. *PrairyErth: (a Deep Map)*. Boston: Houghton Mifflin, 1993.

Horak, Laura. "The Global Distribution of Swedish Silent Film." In *The Blackwell Companion to Nordic Cinema*, edited by Mette Hjort and

Ursula Lindqvist, 457–84. West Sussex: Wiley-Blackwell, 2016.
Kemp, Karen K. "Geographic Information Science and Spatial Analysis for the Humanities." In *The Spatial Humanities: GIS and the Future of Humanities Scholarship*, edited by David J. Bodenhamer, John Corrigan, and Trevor M. Harris, 31–57. Bloomington: Indiana University Press, 2010.
Kessler, Sarah. "MapQuest Founder on the Past, Present and Future of Location Tech." *Mashable*, November 30, 2010. http://mashable.com/2010/11/30/mapquest-founder-interview/.
Klenotic, Jeffrey F. "Space, Place, and the Female Film Exhibitor: The Transformation of Cinema in Small-Town New Hampshire during the 1910s." In *Locating the Moving Image: New Approaches to Film and Place*, edited by Julia Hallam and Les Roberts, 44–79. Bloomington: Indiana University Press, 2014.
Knowles, Anne Kelly. "GIS and History." In *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship*, edited by Anne Kelly Knowles, 1–25. Redlands: ESRI Press, 2008.
———, ed. *Past Time, Past Place: GIS for History*. Redlands: ESRI Press, 2002.
———, ed. *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship*. Redlands: ESRI Press, 2008.
Knowles, Anne Kelly, and Amy Hillier. "Preface." In *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship*, edited by Anne Kelly Knowles and Amy Hillier, xiii–xvii. Redlands: ESRI Press, 2008.
Kraak, Menno-Jan. *Mapping Time: Illustrated by Minard's Map of Napoleon's Russian Campaign of 1812*. Redlands: ESRI Press, 2014.
Lendino, Jamie. "The History of Car GPS Navigation." *PCMAG*, April 16, 2012. http://www.pcmag.com/article2/0,2817,2402755,00.asp.
Loiperdinger, Martin, and Uli Jung, eds. *Importing Asta Nielsen: The International Film Star in the Making, 1910–1914*. KINtop 2. Herts, UK, and Bloomington: John Libbey and Indiana University Press, 2013.
Maltby, Richard, Daniël Biltereyst, and Philippe Meers, eds. *Explorations in New Cinema History: Approaches and Case Studies*. Malden: Wiley-Blackwell, 2011.

Maltby, Richard, Kate Bowles, Deb Verhoeven, and Mike Walsh. *The New Cinema History: A Guide For Researchers*. Oxford: Wiley-Blackwell, 2011.

Marey, Étienne-Jules. *La Méthode graphique dans les sciences expérimentales*. Paris: G. Masson, 1885.

Matless, David. "Book Review: An Atlas of the European Novel 1800–1900." *Progress in Human Geography* 23, no. 4 (December 1, 1999): 659–60. doi:10.1191/030913299677607263.

McBride, Simon, Chris Bellman, and Colin Arrowsmith.. "Developing Space-Time Models and Representation for GIS." In *Proceedings of the Spatial Sciences Coalition Conference*. Canberra, Australia, 2003.

Monmonier, Mark S. *How to Lie with Maps*. Chicago: University of Chicago Press, 1991.

———. *Mapping It Out: Expository Cartography for the Humanities and Social Sciences*. Chicago: University of Chicago Press, 1993.

Moretti, Franco. *Atlas of the European Novel, 1800–1900*. New York: Verso, 1998.

Nowviskie, Bethany. "How to Play with Maps." In *Cultural Mapping and the Digital Sphere: Place and Space*, edited by Ruth Panofsky and Kathleen Kellett, 107–28. Edmonton: University of Alberta Press, 2015.

Sandberg, Mark B. "Location, 'Location': On the Plausibility of Place Substitution." In *Silent Cinema and the Politics of Space*, edited by Jennifer M. Bean, Anupama Kapse, and Laura Horak, 23–46. Bloomington: Indiana University Press, 2014.

Scassa, Teresa, Tracey P. Lauriault, and D. R. Fraser Taylor. "Cybercartography and Traditional Knowledge: Responding to Legal and Ethical Challenges." In *Developments in the Theory and Practice of Cybercartography: Applications and Indigenous Mapping*, edited by D. R. Fraser Taylor and Tracey P. Lauriault, 279–97. Amsterdam and Boston: Elsevier Science, 2014.

Taylor, D. R. Fraser. "Maps and Mapping in the Information Era." In *Proceedings of the 18th International Cartographic Conference*, edited by Lars Ottoson, 1:1–10. Gävle, Sweden, 1997.

———. "Some Recent Developments in the Theory and Practice of Cyber-

cartography: Applications in Indigenous Mapping: An Introduction." In *Developments in the Theory and Practice of Cybercartography Applications and Indigenous Mapping*, edited by D. R. Fraser Taylor and Tracey P. Lauriault, 2–15. Amsterdam and Boston: Elsevier Science, 2014.

Théberge, Paul. "Sound Maps: Music and Sound in Cybercartography." In *Cybercartography — Theory and Practice*, edited by D. R. Fraser Taylor, 4:389–410. Modern Cartography. Amsterdam and Boston: Elsevier Science, 2005.

Thompson, Kristin. *Exporting Entertainment: America in the World Film Market, 1907–34*. London: BFI Pub, 1985.

Tufte, Edward R. *Beautiful Evidence*. Cheshire: Graphics Press, 2006.

———. *Envisioning Information*. Cheshire: Graphics Press, 1990.

———. *The Visual Display of Quantitative Information*. Cheshire: Graphics Press, 1983.

———. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire: Graphics Press, 1997.

Velez-Serna, María. "Mapping Settings and Locations." *Early Cinema in Scotland*, November 30, 2014. http://earlycinema.gla.ac.uk/mapping-settings-and-locations/.

Verhoeven, Deb. "New Cinema History and the Computational Turn." *Beyond Art, Beyond Humanities, Beyond Technology: A New Creativity. Proceedings of the World Congress of Communication and the Arts Conference, University of Minho, Portugal*, 2012.

Verhoeven, Deb, and Colin Arrowsmith. "Mapping the Ill-Disciplined? Spatial Analyses and Historical Change in the Postwar Film Industry." In *Locating the Moving Image: New Approaches to Film and Place*, edited by Julia Hallam and Les Roberts, 106–29. Bloomington: Indiana University Press, 2014.

Verhoeven, Deb, Kate Bowles, and Colin Arrowsmith. "Mapping the Movies: Reflections on the Use of Geospatial Technologies for Historical Cinema Audience Research." In *Digital Tools in Media Studies: Analysis and Research: An Overview*, edited by Michael Ross, Manfred Grauer, and Bernd Freisleben, 69–81. Bielefeld and New Brunswick: Transcript, 2009.

Zwicker, Heather. "Edmonton Pipelines: Living and Playing in the Digital City." In *Cultural Mapping and the Digital Sphere: Place and Space*, edited by Ruth Panofsky and Kathleen Kellett, 129–40. Edmonton: University of Alberta Press, 2015.

FIELD SKETCHES WITH ARCLIGHT: MAPPING THE INDUSTRIAL FILM SECTOR

Kit Hughes

Nontheatrical, institutional, and useful media are marked by geographically dispersed production centers and overlapping and divergent producers, distributors, audiences, exhibition spaces, and genres. This essay puts forth Scaled Entity Search (SES) as a methodology for drawing the broad outlines of institutional media sectors, e.g., industrial film, medical television, or educational media, to provide comparative context difficult to achieve with case studies alone. My title refers to the diagnostic tool long used by geographers to develop productive knowledge about landscapes. In establishing a site's most important features, condensing detail, and determining the interrelation of elements, field sketches allow for broad comparative analysis attuned to change over time and its causes. While we might read the iterative and necessarily piecemeal work of traditional media history scholarship in this same vein, digital tools have the potential to speed up this process dramatically. Arclight, an online application that measures word frequencies in the Media History Digital Library (MHDL) corpus, enables users to map how different sets of entities (film studios, radio stations, stars, genres) trend within the larger media industry ecosystem as it is represented within trade journals and fan and amateur magazines. To demonstrate how SES, the proposed method for using Arclight, might help produce field sketches of the industrial sector, I discuss two avenues of analysis at different scales that move beyond distant-versus-close reading debates.[1] I address institutions, architectures, spaces, and geography by using Arclight to determine what exhibition sites (schools, fairs), physical locations (desks, workspaces), and cities (Chicago, Detroit) appear most frequently in *Business Screen*, the major industrial film trade publication of the twentieth century. I contrast these results with those of other journals, both within the nontheatrical circuit

(*Educational Screen*) and outside of it (*Motion Picture Daily*). I also use Arclight to identify the studios most significant to the trade press and sketch out the rough size and shape of the industry's thousand-plus producers. What follows is preliminary; my goal is not to build a singular representation of the industrial film sector, but to demonstrate how, collectively, scholars might use Arclight to understand large fields of media practice better and at scale. In addition to modeling digital methods and pointing to potentially productive avenues of research, I explore the limitations of using computational analysis to build our knowledge of useful media's many modes in the twentieth century.

CREATING ENTITY LISTS AND BUILDING THE CORPUS

Building entity lists—groups of search terms—can be the most time-consuming and challenging part of the SES process. For this study, I developed four lists designed to provide a broad overview of the field of industrial film—its major producers, geographies, institutions, and physical locations. The first of these was developed from *Business Screen*'s National Survey of Film Production Resources and includes every US-based industrial film production studio exhibited in the journal's listings from 1951, the inaugural year of the survey, to 1974, the final year included in the MHDL.[2] In addition to producer names, *Business Screen*'s annual reviews contain a range of useful information, including executive staff, recent productions, and sponsor clients. According to *Business Screen*, the magazine vetted each company, featuring only those firms able to furnish the above information and describe several productions made over the course of the previous year.[3] When constructing this entity list, I built a spreadsheet that included company name, name variants, founding date, location(s), and the years the firm was included in the review. My company entity list contains 1,344 distinct companies and 135 name variants.

Although the ultimate goal of building entity lists is to run them through Arclight's comparative engine, the process of captur-

ing semistructured data like those found in annual reviews can itself lead to insights.[4] Using the above spreadsheet, for example, I plotted a basic chart describing the rise in industrial production houses included in *Business Screen*'s annual reviews. Figure 1 indicates a sharp rise in production houses between 1954 and 1958 and in 1971, as well as a more general upward trend until about 1972. While these results are not conclusive—more analysis is needed to see whether, on average, the year a given company was first included in the review corresponds with the year it claims as its founding date—they nevertheless provide some guidance as to the rough numbers of active producers participating within the ecosystem of the trade press. Other work might feed the producer locations into geographic information systems (GIS) software to complicate this simplistic growth narrative with details regarding the specific locations of growth and contraction within the industrial film production industry.

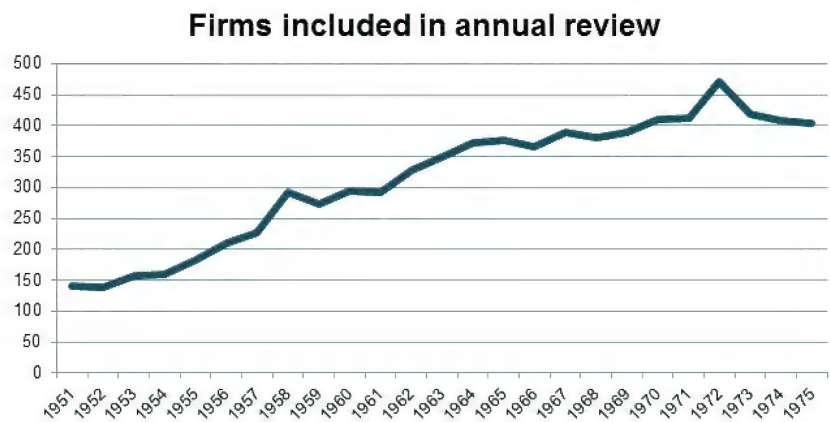


Figure 1. Firms included in *Business Screen*'s annual production review.

My second entity list is itself geographical and contains 311 cities culled from census records and the pages of *Business Screen*. Although the primary dates of my study are 1930 to 1965, my list includes the 100 most populous American cities from 1900 to 1980

(160 cities total) in order to broaden the possibility of locating significant activities in smaller but growing midcentury cities. I supplemented these entities with any additional cities mentioned in *Business Screen*'s annual reviews, e.g., Hollywood, to better tailor the list to my research questions.

My third and fourth lists were attuned to space and location. One compiled the most common nontheatrical exhibition sites, such as schools, workplaces, fairs, and museums. The other gathered terms relating to the physical location of screen technologies within nontheatrical spaces, e.g., desk, shelf, counter. To generate these lists, I consulted prominent anthologies and books within the useful media subfield (*Films that Work*, *Useful Cinema*, *Learning with the Lights Off*, *Ambient Television*, *The Field Guide to Sponsored Films*, and *Medical Visions*), looking for terms already identified by researchers as relevant to the spaces and locations of nontheatrical cinema.[5] I used my general familiarity with *Business Screen* to supplement these lists and tailor my entities focus on industrial media. Given the generic nature of space terms, I consulted a thesaurus to increase my list of location entities from 30 entries to 51 entries. My sites list comprises 108 terms.

The time investment required to compile such lists is better measured in weeks than days. In particular, collecting data from *Business Screen*'s annual reviews—especially as the total number of firms included grew, and companies shuttered, moved, and changed names—was tedious and time consuming. While this represents a potential barrier to scholars interested in using Arclight for scaled projects employing lengthy entity lists, it also suggests opportunities for collaboration. As more people build, use, adapt, and share entity lists tailored to different arenas of research (e.g., silent film, female talent, the studio system), scaled digital analysis will become more accessible to those who cannot afford a major upfront time investment when results are uncertain.

Collaboration is also imperative for evaluating and sharpening the quality of our entity lists. In the context of SES, entity list inclusions and exclusions embed assumptions in digital data collection. On the first count, we might consider how the lists I've chosen for my "field sketch" have distinctly spatial and industrial biases. Even though, as Anna McCarthy notes, space and location have long been used to classify the presumed identities of media consumers, I could have developed lists based more squarely on audiences, with entities such as "school children," "worker," and "farmer."[6] Likewise, I could have followed Thomas Elsaesser's suggestion to attend to the three As: (translated to English) sponsor, occasion, and audience.[7] These alternatives provide worthy avenues for further research.

Last, disambiguation—distinguishing visually identical but semantically different words, e.g., the noun and verb forms of drive—presents structural problems for linguistic analysis. Due to the results' unreliability, I struck words from all four lists that 1) harbored multiple meanings where 2) the less relevant meaning (Eugene as a man's name rather than Eugene, Oregon) was likely to be prominent. Although this hampers the reliability of results—in some cases more seriously than others—Arclight's computational analysis of numerical data to quantify discussions and emphases within the trade press is not meant to stand alone as rigorous, empirical fact. Instead, it offers a way of playfully deforming historical traces and refiguring our approach to research questions. Furthermore, manual disambiguation strategies can help reduce false positives. For example, "library" presents significant problems for an analysis of exhibition given the primacy of film and music libraries in distribution. However, because Arclight allows users to search exact strings of words, one can better target the exhibition-related uses of "library" with terms like "memorial library" and "free library."[8] Potentially even more useful—since it is difficult to cobble together all of the different terms applied to book libraries—is the option to run false

positive threats, like "film library" and "transcription library," to see what proportion of total "library" results they represent. To reduce error further, one can click through Arclight's graphs to access the Lantern results for "library" in order to identify common but unanticipated occurrences of the term. In doing this, I found "Library of Congress" and "Media History Digital Library"—both included in scanned journals' credits pages—also contributed to inflated "library" results. With a list of seventeen terms developed through this process, I was able to identify between 25% and 60% (varied by year) of "library" results as false positives and remove them from my analysis. While some terms will never work due to their inherent ambiguity, routine, iterative processing of terms and results can help overcome the application's limitations for many users.[9]

Although the MHDL provides users with a built corpus, by allowing users to engineer their own corpora within those bounds, Arclight allows for more granular and comparative modes of analysis while emphasizing the impact corpus design makes on SES results. For this field sketch, I ran three of my entity lists (locations, sites, and cities) through six separate journals as well as the agglomerated data from the entire MHDL corpus, so I could compare *Business Screen* results with those from other trade magazines. I chose journals based on their availability spanning 1930 to 1965 and their ability to stand in for a particular genre of trade magazine. My six corpora are *Business Screen* (MHDL holdings: 1938–74), *Educational Screen* (1922–62), *Broadcasting* (1931–56), *Modern Screen* (1930–60), *Motion Picture Daily* (1931–60), and *The Journal of the Society of Motion [and Television] Picture Engineers* (1916–54).

FILLING IN THE SKETCH: EXHIBITION SITES

While Arclight is very good at identifying top trending entities—a potential shorthand for significance—it is perhaps more useful for granular comparative analysis based on the shifting intensities of

term significance over time and across different journals. Taking my sites results as an example, I will tackle both of these modes, beginning with the changing intensity of terms in *Business Screen* over time.

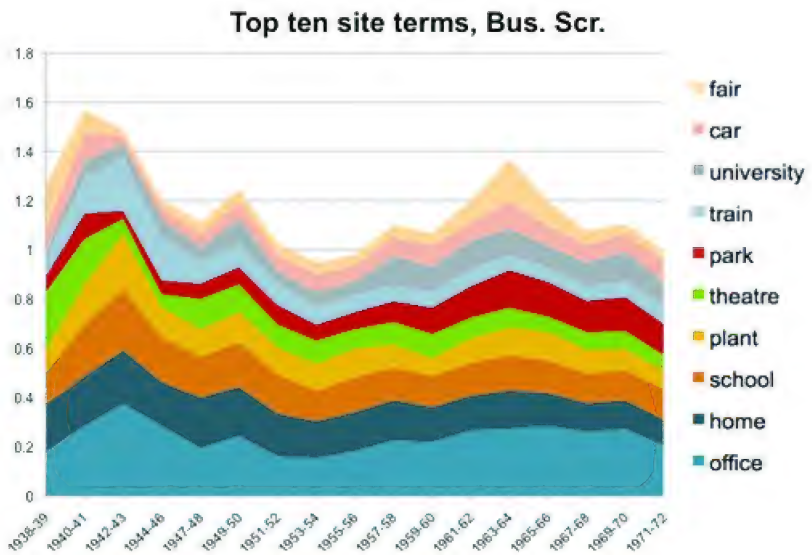


Figure 2. Top ten *Business Screen* site terms returned by Arclight query.

Figure 2 features the top 10 terms within *Business Screen* from 1938 to 1964. Even these limited results provide some insight into both the journal's discussions of space and the limitations of word frequency analyses. First, while I expected "office" to take the top spot in *Business Screen* results, "home" is somewhat higher than I imagined, surpassing "school" and "plant," both key spaces of industrial film practice. A return to the MHDL's full-page scans of the magazine suggests several factors contributed to the prominence of "home": metaphorical invocations of home in advertisements, discussions of professional home-building organizations, advertisements for mobile projection technologies, references to the home office (read: headquarters), mentions of the home

front during war, talk of home economics, and analysis of the home market.[10] Although incomplete, this list both speaks to the varied ways in which notions of home enter the industrial film imaginary and points to the additional work required to fully appreciate the meaning of even a single term as it journeys through a trade press. Close readings of *Business Screen* could more precisely determine how home tends to appear—in what contexts, in whose words, and in what numbers.

Second, the appearance of "car" in *Business Screen*'s top 10 speaks to problems with disambiguation. Included to capture discussions of mobile radio, "car" is far higher than any such mentions would warrant. A Lantern search suggests "car" is used most frequently to refer to the commodity, with some mentions of traffic school and unrelated hyphenates (the "car" of cartridge, caring, carbon and other terms split across two lines) rounding out the results.[11] Although I knew terms like car and train—both products of industries at the forefront of industrial motion picture production—would create problems with disambiguation, I included them in this preliminary study to appeal to comprehensiveness and to measure their exact effect on results. Striking both terms brings "hotel" and "convention" into the top 10—somewhat interrelated terms that bring with them their own challenges, e.g., "convention," "fair," and several other site terms are both locations and occasions. The precise impact of this conflation I leave for other researchers; here it stands as yet another reminder that Arclight users must be cognizant of the way language's complexity challenges word frequency analysis.

For analyses that emphasized broad-based comparisons over top results, I turned to a rudimentary but accessible visualization software. Applying Excel's conditional formatting features to comma-separated values (CSV) files downloaded from Arclight, I created heat maps of my results that used color gradation to indicate cell value. These charts enabled me to identify and quantify hot spots

of activity over time and across results, comparing more information than Arclight's visualizations could meaningfully arrange.

Sample Heat Maps: Cities Data

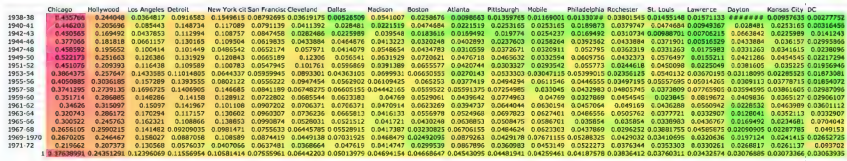


Figure 3. Heat map of top 20 cities mentioned in *Business Screen*. Red indicates highest values; color variation indicates rise and fall in city name prominence.

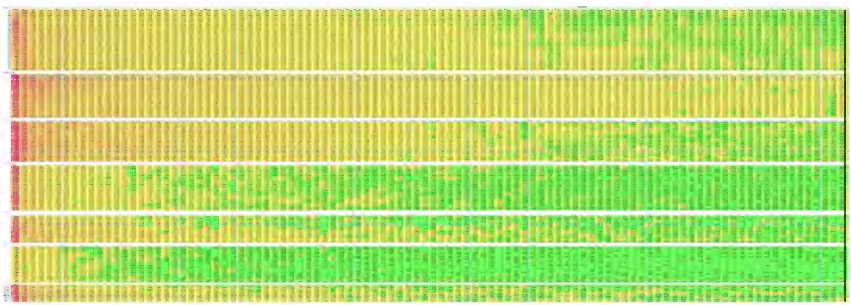


Figure 4. Stacked city heat maps, from top: all journals, *Broadcasting, Motion Picture Daily, Modern Screen, Educational Screen, Journal of the SMPE, Business Screen.*

As with many digital methods, initial results provide as many questions as answers. While "office" remains relatively steady over time, "home" gradually loses ground after 1950. "Classroom," "class," and "school" also dip in the late 1940s and 1950s, though "university" rises steadily. Could the fall in home and school terms indicate increasing specialization in the industrial film sector as filmmakers turn more toward worker audiences? Along these lines, might the uptick in "university" have something to do with vocational training? Or, could we read these same trends as an indication of increased specialization in the trade press itself? If either of the above hypotheses is true, what could explain decreases in "factory?" Could it be that sales industries, like insurance, became more important to the readership of *Busi-*

ness Screen over time? How might analyzing other terms related to school-aged audiences—names of youth organizations like the scouts, for example—complicate our understanding of the changing audiences of industrial film? Shifting to a different set of questions, several terms' low rankings result from their late growth within the temporal bounds of my study. Could rises in words like "sports arena" and "stadium" signal an intensifying relationship between industrial and commercial sports media? As others and I have argued elsewhere, content analysis offers a powerful complement to SES that could help pursue these strands of research.[12] Although these questions are preliminary, they nevertheless point to scales of analysis encouraged by digital tools, as well as these tools' utility for identifying moments of change that can help us better understand the dynamism of the nontheatrical filmmaking industry and one of its major trade papers.

Also revealed by Excel's conditional formatting are hot spots of isolated activity that may express rapid shifts in coverage. A 1940 to 1948 rise in "branches" could signify the growth of franchises as a particular form of industrial organization or the increasing significance of certain types of industries to *Business Screen*, e.g., banking. However, it is perhaps more likely that the increased emphasis on branches is tied to the nontheatrical sector's massive involvement in World War II. To understand this phenomenon better, I returned to page views in Lantern to compare *across* the several journals in my corpus. Cross-journal comparison also helps identify other terms that rise and fall not due to the narrow interests of journals, but in response to larger cultural factors. Indeed, the corpus of all combined journals indicates a small jump in mentions of "branches" during the war and in the immediate postwar era, while *Educational Screen*, the other journal most likely to trade directly in training and education materials, exhibits the most profound peak for "branches" during the war period. A series of Lantern searches confirmed the impact of the war on the increase of "branches" in *Business Screen*.

For those interested in pursuing questions regarding journals' involvement in the war effort, comparing the heat maps for all seven sets of journals helps indicate other "war" terms hidden by relatively static frequencies in *Business Screen*. Stacking journal results on one page results in a visualization (see fig. 4 for an example) similar to those used to compare massive genome datasets, a technique the UW Graphics Lab has recently applied to the ana-lysis of text via TextDNA. The broad scale of this type of visualization allows for easy identification and comparison of terms that spike during war years (branches, camp, canteen, office, plane, theater), and begins to provide a vocabulary for further WWII data work. Likewise, Lantern searches identifying how terms are commonly deployed in context- and topic-specific resources provide additional avenues of entity development (European theater, Mediterranean theater, Pacific theater, China Burma India theater, Western theater, Eastern theater, theater of operations). A researcher could deploy these entities across their own tailored corpora in order to understand how different subfields within the trade press responded to the war and to what intensities. Given the significance of World War II in legitimizing the use of 16mm as a tool of indoctrination and training, questions regarding the role of journals like *Business Screen* in promoting the moving image as a solution to wartime problems are not tangential but central to the mapping of the field in the twentieth century.[13]

Comparison enabled by stacked heat maps also helps explain differences and degrees of specialization across presses. While journals generally share top results (school, home, office, theater), the intensity or centrality of these terms within magazines varies dramatically. The prevalence of "school" in *Educational Screen*, "theatre" in *Motion Picture Daily*, and "home" in *Modern Screen* is not unexpected. However, comparative visualizations help quantify the degrees to which each journal attended to different

spaces. “School” represents an average of 0.575% of all words in the *Educational Screen* corpus—over twice the average of *Business Screen*’s most frequent term, “office” (0.232%). Furthermore, when adding the top four entities for *Educational Screen* (school, university, classroom, and class), these terms comprise almost 1.3% of the magazine’s total word usage, suggesting that discussion in *Educational Screen* may be highly specialized and spatially directed, even in comparison to other nontheatrical magazines like *Business Screen*. That said, differentiation problems dissemble the relative uniqueness of “office” to the latter journal. While “office” ranks high for all journals, beyond the pages of

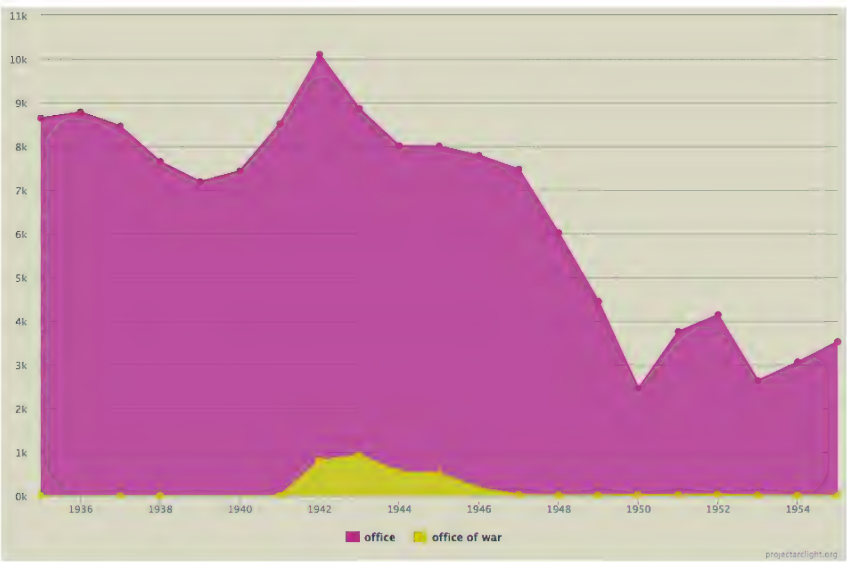


Figure 5. Results for all journals. The yellow peak (“office of war”) suggests that a significant portion of the early 1940s peak in “office” results from war mentions.

Business Screen, it generally refers most frequently to “box office” or “home office” rather than specific kinds of corporate spaces. In *Motion Picture Daily*, for example, over 72% of “office” hits fit into the former category, compared to 7% in *Business Screen*. It may also be that exhibition location is not the best way to dif-

ferentiate between journals; the SMPE journal's lack of attention to exhibition sites both emphasizes its differences from the other journals and points to difficulties that arise when an entity set is not particularly prevalent within a corpus. Since I purposefully chose an array of journals to represent different subfields and my entities concern spaces of exhibition, rather than, for example, different media technologies, these differentiation results are both dramatic and blunt. A study zeroing in on journals that focused on a single area of the industry, e.g., commercial film entertainment, might find more surprising and granular results.

SPACES OF CINEMA

Following much the same process of toggling back and forth from Arclight to Lantern and the MHDL, I examined the data derived from my two other geo-spatial entity lists. Given the introductory goals of this field sketch, I will touch on a few findings and further questions in the hope that others will use this data for their own research projects. Beginning with the spatial data, it is clear that different sets of entities—especially those that are not proper nouns—build different amounts of ambiguity into results. Initially, I hoped that comparing terms like table, desk, and window might help sketch an idea of how media came to be incorporated into cultural-spatial projects by locating projectors, screens, and other equipment within their precise physical surroundings. Indeed, looking at the results for all journals combined suggests that such an analysis may be possible. To take one example, "point-of-purchase" begins to climb steadily in 1951 accompanied by a rise in "check-out" in 1959. Meanwhile, "cosmetic counter" experiences popularity from 1941 to 1953, and "store window" declines in the early 1950s. It is tempting to read this cluster as an indication of shifts in media use in retail space. Did the novel and intimate spectacle offered by small screens lead retailers to shift their presentational address from the store window to the sales floor?[14] Did retailers become more interested in demonstration films tied to specific products?[15] However, these questions rest on

significant assumptions (e.g., that "cosmetic counter" or "checkout" refer in any significant degree to mediated spaces). In addition to the disambiguation problems incurred by generic terms, the frequency numbers for the spatial entities were generally far smaller than their exhibition counterparts.[16] This could mean that the generic materiality of space is discussed less than institutional locations or my entity lists are deficient with regards to how the trades tend to talk about space. As noted above, term frequency analysis promises to be most fruitful when entity lists are tailored to the strengths of a given corpus (hence the preference for Arclight over a program like Ngram Viewer for the analysis of media history). However, when using highly generic terms, even the MHDL may be too broad a corpus for meaningful results.

Although the above questions about the shifting spaces of mediated retail are fairly meaningless—based as they are on overly ambiguous data—Arclight's capabilities to query specific journals can mitigate some of these problems. The results for "table" are instructive. Unlike several other magazines that deployed "table" in highly mixed contexts (dinner table, data table), *Business Screen*'s semi-regular focus on corporate exhibition facilities means "table" appears fairly often as an appendage to projection.[17] Since I focused primarily on *Business Screen* when constructing my entity list, it is unsurprising that its results would speak most to my aims. Using Lantern to add granularity to the "table" results, I developed an additional, partial list of nine terms related to the table as an exhibition technology (fig. 6). Though only a small portion of all *Business Screen* "table" results, this entity subseries suggests that flexibly projected, small-screen devices become increasingly important to the readers (or advertisers) of *Business Screen* in the early-to-mid-1950s. Likewise, "desk," a term geared even more specifically to the architectures of work, peaked a few years later (1957–72) and brought with it its own subset of terms related to small screen projection. To different degrees, the terms "desktop," "desk top," "your desk," "prospects desk," "man-

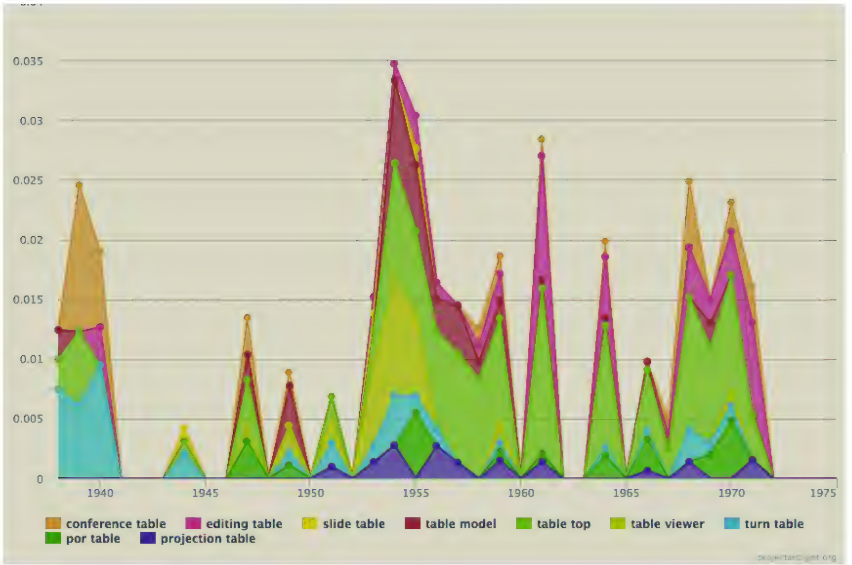


Figure 6. *Business Screen* table results.[21]

agers desk," "across the desk," and "his desk" appear alongside the many portable devices promoted to *Business Screen* readers in the 1950s to 1970s.[18] However, a follow-up search for the term "portable" suggests that mobile technologies and modes of exhibition were central to *Business Screen* well before the prominence of "desktop" screens and machines. Indeed, Haidee Wasson has already located the 1939 World's Fair as a "watershed moment" for small, portable screens used for merchandising and sales.[19] Perhaps the later rise of table and, especially, desk points to an increased articulation of portable projectors to office work where they function as particular types of workspace interfaces. While the MHDL suggests such examples range from the incorporation of a suitcase projector in a traveling salesman's pitch to desktop CCTV controls that shape workplace relations via centralized, mediated management, additional analysis is necessary to trace this articulation.[20] Accounting for only a fraction of the spatial data, these examples suggest that while they are slippery, generic terms may be useful for identifying broad avenues for further research,

provided users tailor their terms tightly to a narrow corpus and engage in significant post-processing of results.

INDUSTRIAL FILM'S FAVORED CITIES

More useful and certain are SES results based on proper nouns. Though not perfect (as indicated by the appearance of Madison, Mobile, and Lawrence in the below chart), these entities experience far fewer problems with disambiguation.[22] Chicago, unsurprisingly, wins the top spot for cities mentioned in *Business Screen*—it is, after all, the magazine's hometown. Long a media capital, the Windy City's prominence within nontheatrical filmmaking is described by Devon Orgeron, Marsha Orgeron and Dan Streible in their recent anthology on educational film.[23] However, the utility of SES comes not with identifying the most important of a set of entities—often already apparent to subject researchers—but in providing qualifying (though quantified) information that places entities within a comparative ecosystem. The chart below not only confirms Chicago's pride of place, it indicates just how dominant Chicago was within the industrial film sector between 1938 and 1972, averaging 35% higher frequency levels than the next-ranked city, as well as 72% and 88% higher frequency than the fifth and tenth top cities, respectively. (The broad strokes of this distribution remain similar whether or not one adds Madison mentions to New York City, bumping the Big Apple to the third spot and pushing Dallas onto the chart in last place).

The yearly data furnished by Arclight provides an even better mechanism by which to contextualize cities' prominence. Figure 4 depicts the top 100 cities across each journal, indicating the degree of geographical centralization—at least in terms of their industrial imaginaries, if not the physical locations of producers and consumers—within each subfield. Given the distributed nature of networks, a high number of locations feature prominently in *Broadcasting*. Diverse cities are far less important to *SMPE*, which exhibits only a handful of locations at high frequency

1	Chicago	.376	11	Atlanta	.045
2	Hollywood	.244	12	Pittsburgh	.045
3	Los Angeles	.124	13	Mobile	.043
4	Detroit	.116	14	Philadelphia	.042
5	New York City	.106	15	Rochester	.038
6	San Francisco	.076	16	St. Louis	.038
7	Cleveland	.064	17	Lawrence	.034
8	Dallas	.050	18	Dayton	.031
9	Madison	.047	19	Kansas City	.031
10	Boston	.047	20	Washington, DC	.031

Table 1. Top 20 cities in *Business Screen*, 1938–72, averaged frequency.

before falling off precipitously. *Modern Screen* showcases similar results, with Hollywood providing a distinct anchor for its coverage of celebrity. *Business Screen*—in a close parallel to *Educational Screen*—falls between these extremes. The industrial journal sees significant frequencies for about 75 different cities, with slightly lower frequency mentions extending to varied cities well beyond the journal's top 100. While Chicago is indisputably central to any understanding of the industrial film sector, these results suggests the value of exploring the industry's relationship to a wide number of cities (and helps identify the likeliest candidates). Furthermore, Chicago's dominance is not static but begins to decline in the late 1950s, around the same time Hollywood and Los Angeles grow in importance. Steady city frequencies are rare in *Business*

Screen, with most cities experiencing peaks and troughs of activity across the short span of some thirty years. Additional work in Lantern and the MHDL could determine whether and how such shifts might be related to create a better understanding of the dynamism of the industrial sector.

Further analysis can also build regional knowledge of the industry that is difficult to glean simply by reading *Business Screen*. A researcher could compare Midwestern cities to the coasts, build city comparison clusters according to local commodity production, or create a snapshot of industrial filmmaking by state. A chart of major Ohio cities, for example, displays the staccato pattern of prominence endemic to many cities and points out potentially meaningful patterns and transitions. What might explain Dayton's rather dramatic decline in the 1960s? Are Cleveland and Cincinnati's twin peaks in 1949 related? How might the health of a state's industry affect its film producers? Are states' borders the most useful way to define regional orientation?

Although the alchemy of what leads to each city's rise and fall differs to some degree, certain factors may contribute similarly to station prominence. Taking Cincinnati's sharp rise between 1955 and 1956 as an example, it becomes clear via Lantern that the city's rising representation is due to editorial content (stories on Cincinnati-based production companies, the release of new sponsored films, industrial personnel briefs, and advertisements featuring Cincinnati sponsors as clients).[24] Although advertising occupied a majority (56%) of Cincinnati hits in 1956, this represents a significant decrease from 1955 (70%). Page-level analysis of Cincinnati hits in Lantern and additional research into the city's industrial (and industrial film) history could further contextualize this shift. Additional case studies—in Ohio and beyond—could likewise situate the Rust Belt city's experience as majority or marginal. Taking full advantage of the comparative breadth enabled by Arclight means building scores of such case studies that could

be set in relation to one another. While the daunting scale of such a project demands collaboration, Arclight results provide a scaffold to help researchers carefully select and situate their case studies.[25]

INDUSTRIAL FILM PRODUCERS

This final section sketches the field from the perspective of production. Although few surprises appear in the top results, like the cities data, the utility comes not necessarily from identifying front-runners, but from more acutely qualifying entities' relationships to one another. The table below, which lists the top results alongside their frequency, founding date, and headquarters city, suggests the most frequently discussed companies (averaged across the entire 1938–74 MHDL run of *Business Screen*) were established early (seven before *Business Screen*) and located in a top five industrial film city. Although Jam Handy's peak position is not unexpected, frequency numbers reveal just how central the firm is to *Business Screen*.[26] While these results help identify important companies and position them within a larger field, my primary interest in production data lies with the hundreds of companies ranked nowhere near these top entities.

Although *Business Screen*'s annual reviews often list hundreds of firms, many of these companies are invisible outside of the reviews' pages. Out of almost 1,500 featured producers, only about 300 companies received any significant coverage in the trade press (fig. 7). Indeed, only 127 companies received 50 or more mentions in *Business Screen* from 1938 to 1972. Nine hundred twenty-two companies received fewer than a dozen mentions in the entire corpus. For the 603 companies that received five or fewer mentions, the annual reviews might represent their only coverage. Whether the short life of so many companies is due to failure or consolidation requires more analysis. Although mentions in *Business Screen* cannot stand as an index to companies' importance within the industrial film sector, as the field's primary

1	Jam Handy Organization	.042	1917	Detroit
2	Audio Productions	.028	1933	New York City
3	Wilding Picture Productions	.020	1914	Chicago
4	Caravel Films	.020	1921	New York City
5	Sarra, Inc.	.019	1937	Chicago
6	Sound Masters	.013	1937	New York City
7	MPO Productions	.012	1946	New York City
8	Parthenon Pictures	.012	1954	Hollywood
9	Atlas Film Corporation	.012	1913	Chicago
10	Dynamic Films, Inc.	.011	1945	New York City

Table 2. Top 10 US industrial film studios in *Business Screen* by page frequency.

journal, it is unlikely that major companies would be absent from its coverage.

Figure 8 further suggests that business film production—as represented by *Business Screen*—was a fickle industry populated by numerous but short-lived companies. Its scatter plot depicts the total number of years a company received coverage in *Business Screen*, organized by total number of overall page hits. The bottom half of companies appeared almost entirety in fewer than five years of the magazine (as did a significant number of more highly ranked companies), emphasizing the volatility of the industry. The sometimes wide variability between a firm's frequency ranking and the number of years it appeared in *Business Screen* suggests

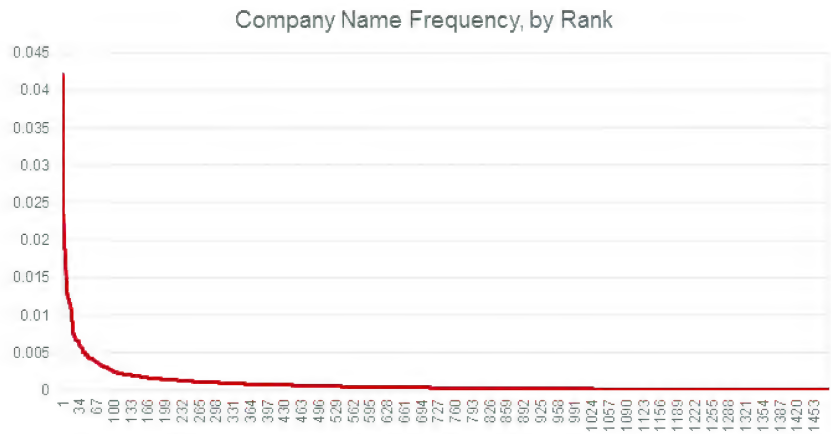


Figure 7. Industrial film studio names, ranked by frequency in *Business Screen*.

that companies built their reputations through different means, pointing to potential case studies for further analysis along these lines. For example, Filmex (in green) was only featured in *Business Screen* for eight years, but ranks sixty-sixth out of all companies. While additional research could determine why Filmex received so much coverage despite its short affiliation with the journal—and how other companies found themselves in different positions across the plot—more ambitious projects could coordinate between Arclight, Lantern, the MHDL, and GIS systems to link company data with cities results for a more distant but still nuanced view of industry dynamics.

Given the close relationship between the industrial and educational film sectors, I also explored companies' frequency results within *Educational Screen* in order to get a quantitative sense of their overlap. Surprisingly, only 211 companies listed in *Business Screen*'s annual reviews made it to the pages of the educational trade, and only about 40 of those companies were mentioned to any significant degree.

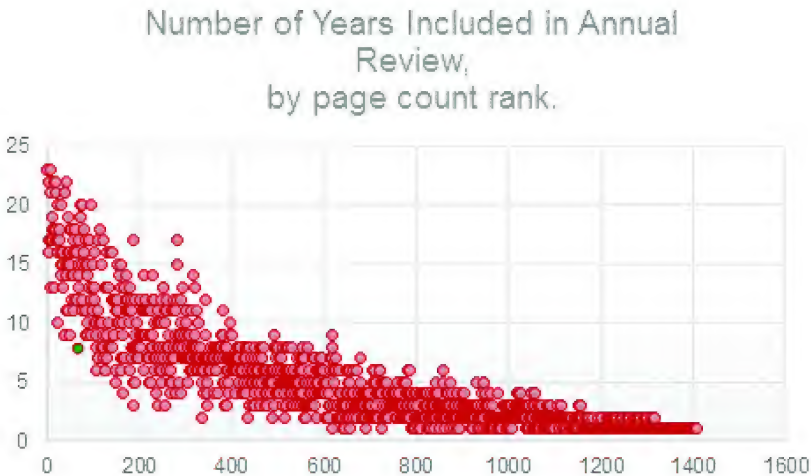


Figure 8. Industrial film studios arranged by number of years included in *Business Screen*'s annual production review (Y axis) and page count frequency (X axis).

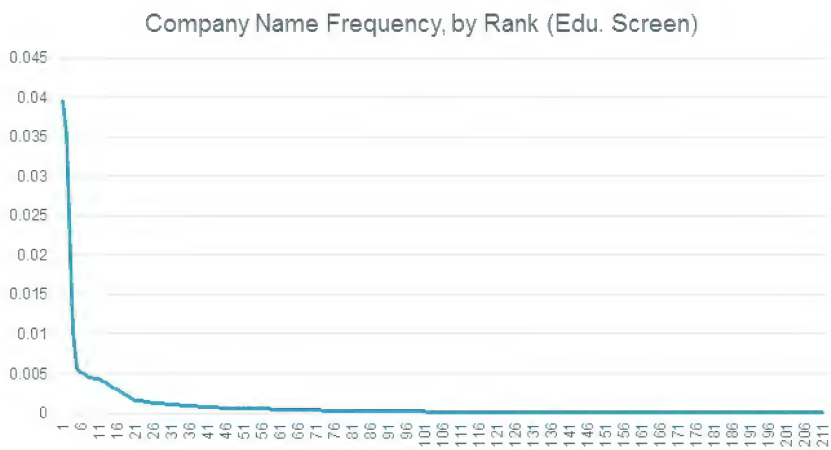


Figure 9. Industrial film studio names, ranked by frequency in *Educational Screen*.

While the low frequency numbers for production companies could arise from a mismatch between entities and corpus due to, for example, a readership more interested in pedagogy and film guides than industrial studios, they may also signify the limits to nontheatrical overlaps in production. The top 10 list for *Educational Screen*, which includes the rank of each company within the

Business Screen results (below), further suggests that the relationship between the two sectors—and, in particular, their trade journals—is not entirely straightforward. Expanding the corpus to include other nontheatrical MHDL journals would doubtless complicate these findings further.

1	Society for Visual Education	.040	1919	Chicago	19
2	United World Films	.010	1946	New York City	17
3	Bray Studios	.006	1914	New York City	68
4	The Jam Handy Organization	.005	1917	Detroit	1
5	International Film Foundation	.005	1945	New York City	289
6	The Princeton Film Center	.004	1941	Princeton	53
7	Films of the Nation	.004	1945	New York City	150
8	The March of Time	.004	1935	New York City	106
9	Paul Hoefler Productions	.002	1939	Hollywood	568
10	Filmfax Productions	.002	1939	New York City	80

Table 3. Top 10 US industrial film studios in *Educational Screen*, by frequency. *Business Screen* rank in last column.

CONCLUSION

Ultimately, one of the difficulties with creating a field sketch is that the territory of useful media history tends to be anticategorical, shifting, and slippery (not unlike written language, as evidenced above). This explains the value in conceptualizing industrial media as series of *Medienverbunde*—constellations of media and media practices coalescing, at least temporarily, within a specific institution or to achieve a particular project—and underscores the value of rich and deeply contextualized case studies.[27] For this, Arclight is no replacement. However, along with SES, it

can provide an additional framework for placing such case studies in relation to one another. The frameworks developed above attempt to describe the institutions, architectures, and locations of industrial media, as well as the landscape of industrial film production. Equally important are field sketches attuned to sponsors, audiences, and a host of other vectors of media practice, such as distributors and manufacturers. Digital projects such as these, however, easily become overwhelming. To take full advantage of digital tools' capabilities for managing inhuman amounts of data, researchers must push beyond their individual capacities and embrace the collaborative effort at the heart of historiographical endeavor.

ENDNOTES

1 For the inciting incident for these debates, see Franco Moretti, "Conjectures on World Literature," *New Left Review* 1, Jan–Feb 2000.

2 I chose to include production houses listed after the major dates of my study 1) because the review lists some producers for the first time years after their actual founding and 2) to create a full entity list that could be helpful to other researchers not bound by the period of my analysis.

3 See "Keys to the Effective Use of the Film Sponsor's Complete Buyer's Guide: Listing Standards Defined," *Business Screen* 20, no. 1 (1959): 99.

4 For more on analyzing credits and other forms of metadata, see Derek Long, "Excavating Film History with Metadata Analysis: Building and Searching the ECHO Early Cinema Credits Database," in this volume.

5 Vinzenz Hediger and Patrick Vonderau, eds., *Films that Work: Industrial Film and the Productivity of Media* (Amsterdam: Amsterdam University Press, 2009); Charles R. Acland and Haidee Wasson, eds., *Useful Cinema* (Durham: Duke University Press, 2011); Devin Orgeron, Marsha Orgeron, and Dan Streible, eds. *Learning with the Lights Off* (New York: Oxford University Press, 2012); Anna McCarthy, *Ambient Television: Visual Culture and Public Space* (Durham: Duke University Press, 2001); Kirsten Ostherr, *Medical Visions* (New York: Oxford University Press, 2013); Rick

Prelinger, *The Field Guide to Sponsored Films* (San Francisco: National Film Preservation Foundation, 2006).

6 Anna McCarthy, 61, 92.

7 Thomas Elsaesser, "Archives and Archaeologies: The Place of Non-Fiction Film in Contemporary Media," in *Films that Work*, 23.

8 This is easier for terms like "Washington, DC" that have strict and limited synonyms.

9 Perhaps an even more dramatic example is the mid-1950s rise in "temple" in *Business Screen*, which stuck out not least because "church" remains relatively stable over the same time period. A quick Lantern search of the MHDL revealed that 75% of "temple" results occurred as instances of "Temple Street" or "Temple St," which in turn were most likely to appear in the advertised addresses of Parthenon Pictures and the Deseret Book Company.

10 "Home and Film: A Survey," *Business Screen* 15, no. 6 (1954): 31; J. Talcott, "Stauffer Closes '7 Out of 10' Using Slidefilm on Home Calls," *Business Screen* 19, no. 6 (1958): 43–44; Michael J. Ritt, "New Communications Center at Chicago's Combined Insurance," *Business Screen* 33, no. 5 (1972): 27.

11 "Wishes on Wheels," *Business Screen* 16, no. 4 (1955): 1; "General Tire Shows a Quality Line on 8mm," *Business Screen* 24, no. 3 (1963): 37.

12 Kit Hughes, Eric Hoyt, Derek Long, Kevin Ponto, and Tony Tran, "Hacking Radio History's Data: Station Call Signs, Digitized Magazines, and Scaled Entity Search," *Media Industries* 2, no. 2. 2015. http://www.mediaindustriesjournal.org/index.php/mij/article/view/128/182

13 See, Kristen Ostherr, "Health Films, Cold War, and the Production of Patriotic Audiences: *The Body Fights Bacteria* (1949)," in *Useful Cinema*, 104.

14 How might this build on the relations between looking, gender, and consumption described by Anne Friedberg in *Window Shopping: Cinema and the Postmodern* (Berkeley: University of California Press, 1993): 66?

15 See McCarthy, *Ambient Television*, 69–80.

16 For example, "table," the top space result in *Business Screen* at

0.046 frequency, garnered less than a quarter of the mentions by percentage than did "office" (0.232).

17 "Dual-Purpose Screening Room: '48 Model," *Business Screen* 4, no. 9 (1948): 24–25. In addition to one-off features, *Business Screen* also featured at least one recurring column on corporate media architectures. Bob Seymour, "Hubert Wilke to be regular Business Screen columnist," *Business Screen* 33, no. 3 (1972): 8.

18 "This vice-president is making a sound movie while he sits at his desk," *Business Screen* 18, no. 3 (1957): 43; "New Super 8," *Business Screen* 26, no. 8 (1965): 1; "HPI Has New 8mm Wide-Angle Lens for Kodak Automatic 8," *Business Screen* 25, no. 4 (1964): 58.

19 Haidee Wasson, "The Other Small Screen: Moving Images at New York's World Fair, 1939," *Canadian Journal of Film Studies* 21, no. 1 (2012): 93.

20 I address some of these questions using more traditional methods in "Corporate Channels: How American Business and Industry Made Television Useful" (dissertation, University of Wisconsin–Madison, 2015), 191–363. When building entity lists for this purpose, the distinctive brand names of portable projectors would be a good place to start. "In 8mm Sound Field New 'Videotronic 8,'" *Business Screen* 22, no. 6 (1966): 51; "Sight, Sound Desk-Top Communication Center," *Business Screen* 30, no. 12 (1969): 74–75.

21 "Por table" is a hyphenate of "portable."

22 While Madison ranks higher due to the prevalence of media industry addresses on Madison Avenue, mobile appears fairly often in relation to portable exhibition technologies. Lawrence's numbers are inflated due to its confusion with a common first name and Robert Lawrence Productions, a Canadian film studio.

23 Michael Curtin, "Media Capital: Towards the Study of Spatial Flows," *International Journal of Cultural Studies* 6, no. 2 (2003): 206–10; Devin Orgeron, Marsha Orgeron, and Dan Streible, "A History of Learning with the Lights Off," in *Learning with the Lights Off*, 48.

24 "Sterling Names Edel Ad Chief," *Business Screen* 17, no. 7 (1955): 20; "A Spokesman for Advertising," *Business Screen* 17, no. 4 (1955): 40; "Lasky Film Productions, Inc. Appoints Geeding Sales Chief," *Business*

Screen 17, no. 3 (1955): 58; "Ohio 'Select List' Papers Film Their Market Story," *Business Screen* 17, no. 2 (1955): 53; "Cinescript, Cincinnati Offers Industry Script Service," *Business Screen* 17, no. 1 (1955): 165.

25 For a model of this approach, see Kit Hughes et al., "Hacking Radio History's Data," n.p.

26 Frequency numbers are even higher for "Jam Handy" (1,394 hits) than "The Jam Handy Organization" (912). This effect is more dramatic for "Wilding" (1,077) over "Wilding Picture Productions" (357), pushing Wilding to the number two spot in the above chart. "Wilding" also garnered 25 hits in *Educational Screen*—21 more than its more official title. Additional analysis could determine the degree to which this is an issue for other firms and whether companies' full names are more commonly used in specific contexts such as advertising. "IMAGE" has been removed from both sets of results due to disambiguation difficulties. For ambiguous company names, addresses and Boolean searches incorporating operation locations could help further parse results.

27 Elsaesser, 22.

BIBLIOGRAPHY

Acland, Charles R. and Haidee Wasson, eds. *Useful Cinema*. Durham: Duke University Press, 2011.

Curtin, Michael. "Media Capital: Towards the Study of Spatial Flows." *International Journal of Cultural Studies* 6, no. 2 (2003): 202–28.

Elsaesser, Thomas. "Archives and Archaeologies: The Place of Non-Fiction Film in Contemporary Media." In *Films that Work*, edited by Vinzenz Hediger and Patrick Vonderau, 19–34. Amsterdam: Amsterdam University Press, 2009.

Friedberg, Anne. *Window Shopping: Cinema and the Postmodern*. Berkeley: University of California Press, 1993.

Hediger, Vinzenz and Patrick Vonderau, eds. *Films that Work: Industrial Film and the Productivity of Media*. Amsterdam: Amsterdam University Press, 2009.

Hughes, Kit. "Corporate Channels: How American Business and Industry Made Television Useful." Dissertation, University of Wisconsin–Madison, 2015.

Hughes, Kit, Eric Hoyt, Derek Long, Kevin Ponto, and Tony Tran. "Hacking Radio History's Data: Station Call Letter, Digitized Magazines, and Scaled Entity Search." *Media Industries Journal* 2, no. 2 (2015). http://www.mediaindustriesjournal.org/index.php/mij/article/view/128/182

Long, Derek. "Excavating Film History with Metadata Analysis: Building and Searching the ECHO Early Cinema Credits Database." In *The Arclight Guidebook to Media History and the Digital Humanities,* edited by Charles R. Acland and Eric Hoyt. Falmer: REFRAME/Project Arclight, 2016.

McCarthy, Anna. *Ambient Television: Visual Culture and Public Space*. Durham: Duke University Press, 2001.

Moretti, Franco. "Conjectures on World Literature." *New Left Review* 1 (Jan–Feb 2000). http://newleftreview.org/II/1/franco-moretti-conjectures-on-world-literature

Orgeron, Devin, Marsha Orgeron, and Dan Streible, eds. *Learning with the Lights Off*. New York: Oxford University Press, 2012.

Orgeron, Devin, Marsha Orgeron, and Dan Streible. "A History of Learning with the Lights Off." In *Learning with the Lights Off*, edited by Devin Orgeron, Marsha Orgeron, and Dan Streible, 15–66. New York: Oxford University Press, 2012.

Ostherr, Kristen. "Health Films, Cold War, and the Production of Patriotic Audiences: *The Body Fights Bacteria* (1949)." In *Useful Cinema,* edited by Charles R. Acland and Haidee Wasson, 103–22. Durham: Duke University Press, 2011.

———. *Medical Visions: Producing the Patient through Film, Television, and Imaging Technologies*. New York: Oxford University Press, 2013.

Prelinger, Rick. *The Field Guide to Sponsored Films*. San Francisco: National Film Preservation Foundation, 2006.

Wasson, Haidee. "The Other Small Screen: Moving Images at New York's World Fair, 1939." *Canadian Journal of Film Studies* 21, no. 1 (2012): 81–103.

PART II: **APPROACHING THE DATABASE**

LOW-TECH DIGITAL

Charles R. Acland

Many areas of Canadian moving image heritage are victims of neglect. But no neglect is as extreme as that experienced by industrial, educational, and sponsored film, a condition shared by many other countries too. These nonfiction films tend to fall out of nearly everyone's attention, including documentary film historians who have been more attentive to the socially engaged, politically charged, or personal film. Indeed, so little has been written about educational, industrial, and sponsored films that, coupled with the lack of coordinated efforts to preserve these artifacts, they hardly figure in the standard histories of Canadian moving image culture as anything more than an afterthought and a throw-away reference.[1] Moreover, when they are discussed, it is the publicly funded films that receive the lion's share of attention. Most famously, this includes those of the National Film Board of Canada, whose existence rests confidently at the center of most writing about Canadian film, and whose ongoing film digitization efforts provide a presence for at least some of their productions online (www.nfb.ca). And yet, industrial, educational, and sponsored film was by far the most active area from the early years of filmmaking in Canada onward. The number of films produced is, conservatively estimated based on existing production company catalogues, in the tens of thousands, and the number of production companies is in the multiple hundreds. Many of these films circulated to international markets and were released in multiple languages, and works were a mainstay of schools, governments, factories, hospitals, and religious organizations, to name just a few of the institutions that embraced localized screenings. While some production companies lasted only a short time, others were active for decades, most influentially Associated Screen News of Montreal and Crawley Films of Ottawa.

There is a certain urgency to our situation. Many holding institutions have been in the process of discarding celluloid prints of these films, and, as most made before the 1970s were never transferred to videotape let alone DVD or otherwise digitized, this means they are in danger of being lost completely. Many of those production companies or commissioning bodies never understood that the films might have any special historical value, so works were not systematically deposited in stable archival settings, like Library and Archives Canada.

What sort of path forward for moving image history might be forged given this state of affairs? More pointedly, what role might online resources and analytics play? To explore those questions, I offer some considerations of the conceptual field for digital humanities work as well as a description of the Canadian Educational, Sponsored, and Industrial Film Project at Concordia University. My goal in this essay is to advance an argument for simple deployments of digital modes, for "low-tech digital," which not only have advantages in terms of cost and technical skill, but also are essential research components that can be glanced over in the leap to algorithmic abstractions. A crucial question floats, often silently, behind digital humanities work: just because the digital tools can do something, should we use them? When are they but a distraction, or even worse, a hindrance to our process of investigating and intervening in cultural processes? When do we productively find a new perspective onto a subject and when do we end up with graphic nonsense?

To begin, we ought to acknowledge that the ever-lengthening twilight of celluloid has produced a considerable ferment of scholarly production on the minor genres of educational, industrial, and amateur film. This production includes some books—*Useful Cinema*, but the especially notable *Films that Work* and *Learning with the Lights Off*—and also special issues of journals, courses, and conference themes.[2] This work stands on the shoulders of an

earlier vanguard, including contributions from Anthony Slide, Patricia R. Zimmermann, Dan Streible, and Rick Prelinger. Their influence is such that research on nontheatrical film is no longer as slender as it used to be; we are at the point that instructional and industrial film, however marginal in general historical narratives, are not entirely absent genres. If you are a PhD student, there is now a bibliography that can be assembled for your exams and thesis proposal.

This scholarly activity is part of a significant re-direction of research in many cultural domains. This shift in focus has been achieved through developing forms of scholarly legitimation for previously devalued culture and practices, filmic and otherwise, involving materials used in disparate institutional settings: military training, educational curricula, science demonstration, psychological experimentation, medical information, bureaucratic operations, farm promotion, and so on. There is no finer essay that captures this legitimation than John Guillory's "The Memo and Modernity," in which he challenges literary scholars to pay attention to that "great mass of writing that is neither scientific nor literary but exists *primarily* to transmit information."[3] The quintessential example of this information genre is, he writes, the memo, which grew from the rising importance of internal communication to large-scale business and organizational entities in the nineteenth century and to new forms of managerial practice. One of the particularities of the memo is that it is both ephemeral, having a relatively short currency for the information it contains, and permanent, as standard practice dictates that it be filed and kept. Guillory goes on to make some interesting claims about the memo as a product of the collapse of rhetoric. But for my purposes here, consider how instructional and industrial film developed and operated with similar connections to managerial practice and to institutional maintenance. These films likewise functioned as institutionally specific communication vehicles, designed for precise uses and goals. And, consider too how turning to these genres

opens up a "great mass" of filmic texts that we had been content to leave aside, creating a sizeable new corpus that immediately draws us to the concerns and limits of how humanities scholars handle "big data." What does it mean to confront all those films that no one has written about?

Related to this "great mass," we face and, in many ways, have normalized such technological upheavals as large-scale digitization of rare and obscure documents, accessibility to research resources that was unimaginable only a few years ago, new modes for seeking materials, reliance on machine reading, and explorations in computational research among humanities scholars. More to the point, while there are some renewed efforts to hold materials in their original formats, and while many of us continue to rely upon celluloid prints for our research, for ease of access digitization has become a major feature in the way we consider and encounter moving image eras of the past in films, periodicals, books, and more. In the process of building a research domain, previously devalued resources have been made available and accessed. Digitization has added a level of durability (at least this is what we tell ourselves at the moment—forever for now) and ease of access and circulation (theoretically, apart from those pesky financial and technological barriers one habitually encounters). Even if you are not especially algorithmically literate, it is nonetheless highly likely that some aspect of online searching has figured in your work. We are so acclimatized to the availability and, frankly, convenience of web-based print, image, and sound materials that the modifier "digital" no longer specifies in the way it used to. Simply put, with alterations of historical method and of the evidentiary value of newly accessible materials, there is a dove-tailed development in play; alongside a critical and conceptual embrace of minor, though plentiful, "useful" media forms, we have a technological infrastructure that builds a particular approach to these materials.

As with all forms of archival practice, whether conducted by actual archives or the more vernacular para-archival projects, storage and availability generates a *post-useful* phase in the life cycle of an artifact. The memo when saved serves a different institution and function (the academy, history) than the one it was born to advance; the same thing goes for some of the films we now capture in our current studies of instructional and industrial shorts. We can't help but re-direct the historical record, and re-direction it needs. So what exactly are we doing when we re-animate works, especially works that were, by design, quite narrowly conceived and deployed? What happens when the minor or marginal forms, or textual instances or iterations, are available next to the grand resonating and lasting ones? What happens when the ephemeral is no longer such and is given a life well beyond its initial functionality, an after-life as historical evidence? After making far too little of so much, when might we be making too much of so little? I've certainly seen the single ephemeral fragment mobilized by critics and scholars not so much in service of the fullness of the historical record, nor the investigation of concrete ideological or discursive structures, but for the purpose of poetic flights of fancy about "pastness," ruins, and romanticist longing. Is the concept of the ephemeral too damn sexy for its own good?

Conversely, trying to read and investigate the whole of a film corpus situates us in an equally tenuous spot. As we access novel materials and amass works that had previously been destined for landfills, we contribute to the empire of data dirt and to supplemental digital availability. At one level, this is right. And Rick Prelinger, in his archival and writing practice, has given us some of the best reasoning on this count. But there is a curious effect here, where the ephemeral, the minor work, gets resituated as worthy of archival and scholarly attention. The ephemeral grows up, matures, leaves behind its youthful inconsequential ways, becomes serious, and runs through a conversion experience to become . . . data. I think of it as a reversal of the fairy tale where,

in this version, the beautiful ephemeral swan becomes the ugly data duckling. We revalue the mass of marginalia and produce a degree zero of media comparability as digital data. The previously varied existences as paper and celluloid, as mass distributed and organizationally specific, as promotional and instructional, are smoothed out as a consistent and uniform landscape of data. Context is absented, in the process effacing any special contributions to media history that may have been possible on the part of any particular media format and material.

And here is the clincher: we do not always find the abundance that is supposed to be there. The rush to digital methods and analytics presumes the existence of a workable and representative digitized dataset. And the sheer volume of online materials leads many to believe that a point of critical mass has been reached and that we may proceed with studies of metadata and topic modeling. In Matthew Jockers's work, for instance his book *Macroanalysis*, there are plenty of references to "extracting" and "harvesting" evidence. One has the impression that material sits there, waiting to be released, hence the aptness of the "mining" metaphor. Jockers even describes "close mining."[4] The raw integers that are repeated to signal the vastness of a digital corpus—a million pages, a million books, a million entries, etc.—tell us nothing of the completeness, representativeness, and robustness of that collection with respect to the actual historical category in question. The wonderment conjured by big numbers provokes what Lisa Gitelman has called our "false sense of completeness" that we attribute to datasets.[5]

The Canadian Educational, Sponsored, and Industrial Film (CESIF) project is an effort to put a vastly under-documented realm of cultural life appropriately into the historical record, and to do so with attention to context built into the enterprise. I lead this project at Concordia University with Louis Pelletier and the participation of numerous graduate research assistants. With CESIF, we

began with the premise that prior to any interpretive and analytical stage, prior to any critical assessment, a stock-taking exercise was essential. We realized that it was folly to rely on what might at first glance appear to be a deep well of already digitized materials. The existing online holdings of films and records were simply partial, dispersed, without coordination, and without any specific mandate to advance a full understanding of this neglected realm. We have also been committed to making the product of our labor accessible, easy to use, and free of charge.

Our first step in coordinating and prompting research on these films has been the construction of an open-access online database about the films, using Drupal. We mapped the site to be searchable, to be bilingual in French and English, and to offer related research resources. Entries on films are easily organized by subject, title, producer, date, and accession number, which we assign upon the entry's creation. Starting with available online catalogues, followed by confirmation and cross-referencing with company catalogues, the CESIF research team scoured libraries and archives across the country, identifying titles produced by, or coproduced with, private Canadian film companies. The project does not encompass amateur nor public production venues, though public bodies did regularly commission work from private concerns, so those titles would be included in our framework. Each was then given an entry that included such information as film credits, subject matter, and holding institutions. The idea is that this database captures information about and provides a single coordinated hub for this fading cultural domain. The database additionally allows us to identify those works that had a specifically national presence, such that we can identify titles that, for whatever reason, were deemed appropriate for multiple holding institutions to acquire and maintain as part of their collections. It is important to understand that the CESIF project is not just about finding fragments in a dumpster nor appreciation for any particular "orphan" film—though such approaches may follow and

may be valuable illuminations of the corpus—but about offering a context for their presence by seeing them as part of circulating institutions. We currently have more than 4,000 entries completed, and we are moving to the next phase of generating essay content, organizing scholarly events, and embedding videos of digital version of some films. Longer-term goals include a historical survey of the industry and portraits of notable production entities, along with descriptions of and critical commentary on exemplary films. While we are not in a position to archive film prints, we have done some targeted digitization. Importantly, we intend CESIF to help provide a powerful argument for the work that archivists and preservationists do by demonstrating the range of materials of this kind that were produced in Canada, in addition to helping shift scholarly attention to this major under-explored area of Canadian moving image culture.

Exceptionally rare is the single moving image artifact that has a resounding textual or discursive force upon history, and, in fact, in most cases such elevation would skew actual influence. But educational, instructional, and sponsored films decorated the institutional landscape of Canadian life by the thousands and were standard features of organizational operations for decades. Canadian cinema has often been characterized as an absence—of films, of talent, of audiences. CESIF challenges this myth, clearly and materially documenting the robust moving image practice that existed as private film production and distribution companies catered to institutional requirements. We might think of this as a form of "administrative filmmaking," produced as a result of contract and sponsorship, and it surely doesn't satisfy anyone's conventional idea of a national cinema culture. And yet, that common sense idea about "national cinema" reveals more about the legacy relationship that category has with art practices than it does about the actuality of Canadian life. The films in CESIF represent a form of "useful cinema"—filmmaking by commission, by contract, for specific institutional or commercial purposes—casting

film as a technology in the service of existing institutional needs. They were not radical or subversive; they were "new media" that helped modernize institutional operations, creating, if you will, media subjects appropriate to the extension of institutional formations. As documentation of the visual culture of the country, these films capture representational strategies and discursive contestation that were an ordinary part of daily leisure and work environments in Canada.

In our discussions about scholarship, the plenitude of online historical materials, and the methods of deriving meaning from those materials, at times I can't help but hear—and be put off by—a tone of triumphalism sounded by big data enthusiasts. Digital methods promise to ease the labor of handling the new corpus of limitless data, but they do so without accounting for the associated labor of getting up to speed with those methods. They also promise to reveal imperceptible shadings of truth hidden in the deep ocean of digital media artifacts. Here, digital methods can seem like an answer to a question we've not yet posed, as though they are saviors to problems of their own making. Ted Underwood has identified, and is critical of, this ideologically charged "endgame where 'data' finally displaces all 'theory.'"[6] In this vein, it is crucial that as we move forward with our digital scholarship we avoid what visual design historian Edward Tufte calls "chart-junk," the cluttered and incomprehensible graphics that have become pandemic in our era of digital scholarship. We don't need more op art illusions posing as evidence.[7]

CESIF, then, is consciously a relatively simple "low-tech" response. Sure, we can use it as a database, but in the end it is a catalogue. And I want to speak up for and encourage this focus. As our online material enriches and deepens, the curator and the finding aid are going to become ever more crucial to how we navigate these materials. In this respect, I contend that one of the most important sparks to American scholarship on nontheatrical film has been

Rick Prelinger's *Field Guide to Sponsored Films*.[8] Ostensibly an annotated filmography, it operates as a finding aid and a point of consolidation for an emerging research domain. Itself a "low-tech" digital solution, the guide is available for free online. It performs, and has helped others perform, the sort of canon-busting work Guillory calls for in his aforementioned article. Some may see here a parallel move to create a new alternative canon. But, as Matthew Wilkens reminds us, canons have always been a way to deal with abundance.[9] Given the frustrating fact that human existence is stuck in this finite temporal dimension, we have to make choices about where to spend our time. Having mechanisms to guide us are crucial, however productively debatable they continue to be. In the very least, part of what our era of digital archives, digital humanities, and moving image analytics has offered us is a reflective and elevated sense of responsibility for the "memory infrastructure,"[10] and not just to celebrate and ruminate about the random dust of neglected works or to salivate when faced with the vastness of a new realm of fresh textual meat.

The optimism for digital humanities, and data culture more generally, can leave the impression that the Googlization of everything, as Siva Vaidyanathan put it, has resolved Raymond Williams's concerns about "selective tradition."[11] But his apprehension still applies, first because the Googlization of everything is a wish, a dream, and will inevitably produce both predictable and as-yet-unforeseen gaps in the materials available through digital means. Second, "selective tradition" refers to the gravitational pull toward certain forms of historical writing and ways of attending to evidence and away from others, which digital data, however abundant, will not circumvent. Third, access issues and material barriers to those datasets for scholarly work will always arise. And fourth, we cannot ignore the issue of how we end up living with the strains of the past. As Williams put it, "tradition is not just the 'surviving past' but is selectively incorporated into our lives."[12] However big this big data era gets, these four features of determi-

nation (and I'm sure there are others) will remain with us as we attempt to adjust our scholarly methods accordingly. These issues should invigorate our work rather than slow it down. They should encourage us to avoid being bedazzled by algorithmic magic and to get down to some of the basic scholarly labor of taking account of the production and circulation of materials that have consequence upon our lives. Working with exactly these parameters in mind is the only way we are going to get to a deep, contingent understanding of and response to the historical conditions that meaningfully structure the world around us.

ENDNOTES

1 Two recent contributions stand as significant exceptions that are helping develop interest in this area: Zoë Druick and Gerda Cammaer, eds., *Cinephemera: Archives, Ephemeral Cinema, and New Screen Histories in Canada* (Montreal: McGill-Queen's University Press, 2014); Monika Kin Gagnon and Janine Marchessault, eds., *Re/imagining Cinema: Films at Expo '67* (Montreal: McGill-Queen's University Press, 2014).

2 Charles R. Acland and Haidee Wasson, eds., *Useful Cinema* (Durham: Duke University Press, 2011); Vinzenz Hediger and Patrick Vonderau, eds., *Films that Work: Industrial Film and the Productivity of Media* (Amsterdam: Amsterdam University Press, 2009); Devin Orgeron, Marsha Orgeron, and Dan Streible, eds., *Learning with the Lights Off: Educational Film in the United States* (Oxford: Oxford University Press, 2012).

3 John Guillory, "The Memo and Modernity," *Critical Inquiry* 31, no. 1 (2004): 111.

4 Matthew Jockers, *Macroanalysis*, (Urbana: University of Illinois Press, 2013), 171.

5 Lisa Gitelman, "Searching and Thinking about Searching," *Representations* 127, no. 1 (Summer 2014): 78.

6 Ted Underwood, "Theorizing Research Practices We Forgot to Theorize Twenty Years Ago," *Representations* 127, no. 1 (Summer 2014): 69.

7 Edward Tufte, *The Visual Display of Quantitative Information*, (Cheshire: Graphic Press, 1983).

8 Rick Prelinger, *The Field Guide to Sponsored Films*, (San Francisco: National Film Preservation Foundation, 2006).

9 Matthew Wilkens, "Canons, Close Reading, and the Evolution of Method," in *Debates in the Digital Humanities*, ed. Matthew K. Gold (Minneapolis: University of Minnesota Press, 2013): 249–58.

10 Randall Mason, *The Once and Future New York: Historic Preservation and the Modern City*, (Minneapolis: University of Minnesota Press, 2009).

11 Raymond Williams, *Marxism and Literature,* (New York: Oxford University Press, 1977), 116.

12 Ibid., 115.

BIBLIOGRAPHY

Acland, Charles R. and Haidee Wasson, eds. *Useful Cinema*. Durham: Duke University Press, 2011.

Druick, Zoë and Gerda Cammaer, eds. *Cinephemera: Archives, Ephemeral Cinema, and New Screen Histories in Canada*. Montreal: McGill-Queen's University Press, 2014.

Gagnon, Monika Kin and Janine Marchessault, eds. *Re/imagining Cinema: Films at Expo '67*. Montreal: McGill-Queen's University Press, 2014.

Gitelman, Lisa. "Searching and Thinking about Searching." *Representations* 127, no. 1 (Summer 2014): 73–82.

Guillory, John. "The Memo and Modernity." *Critical Inquiry* 31, no. 1 (2004): 108–31.

Hediger, Vinzenz and Patrick Vonderau, eds. *Films that Work: Industrial Film and the Productivity of Media*. Amsterdam: Amsterdam University Press, 2009.

Jockers, Matthew. *Macroanalysis*. Urbana: University of Illinois Press, 2013.

Mason, Randall. *The Once and Future New York: Historic Preservation and the Modern City*. Minneapolis: University of Minnesota Press, 2009.

Orgeron, Devin, Marsha Orgeron, and Dan Streible, eds. *Learning with the Lights Off: Educational Film in the United States*. Oxford: Oxford University Press, 2012.

Prelinger, Rick. *The Field Guide to Sponsored Films*. San Francisco: Na-

tional Film Preservation Foundation, 2006.
Tufte, Edward. *The Visual Display of Quantitative Information*. Cheshire: Graphic Press, 1983.
Underwood, Ted. "Theorizing Research Practices We Forgot to Theorize Twenty Years Ago." *Representations* 127, no. 1 (Summer 2014): 64–72.
Vaidyanathan, Siva. *The Googlization of Everything (And Why We Should Worry)*. Berkeley: University of California Press, 2011.
Wilkens, Matthew. "Canons, Close Reading, and the Evolution of Method." In *Debates in the Digital Humanities*, ed. Matthew K. Gold. Minneapolis: University of Minnesota Press, 2013, 249–58.
Williams, Raymond. *Marxism and Literature*. New York: Oxford University Press, 1977.

EXCAVATING FILM HISTORY WITH METADATA ANALYSIS: BUILDING AND SEARCHING THE ECHO EARLY CINEMA CREDITS DATABASE

Derek Long

INTRODUCTION

Media scholars may think of metadata—data about data—as a new subject of inquiry, one intimately connected with the so-called "digital turn" in media studies.[1] But consciously or not, film and media scholars have always studied, used, and perused metadata. One of the most important purposes of metadata (as defined in library science) is resource discovery, and film and media studies scholars use metadata whenever they seek to discover broader connections and patterns of meaning that go beyond individual works—that is, nearly all the time.[2] In film studies, for instance, genre criticism is profoundly dependent on descriptive metadata, since it examines particular films alongside contemporaries of the same genre.[3] The same paradigm is true for scholars making claims about authorship, who must naturally know the exact films a particular director, writer, cinematographer, or star worked on. Film analysis often makes extensive use of both descriptive and structural metadata, because films are extraordinarily, if not supremely, replete with images whose formal aspects can be described using metadata. A formal analytical study of a scene, sequence, or even an entire film will inevitably start with what is fundamentally an exercise in metadata collection, since the act of careful description that starts any rigorous analysis requires metadata on salient formal aspects of the film.[4] Many of the contributions in this collection persuasively demonstrate the power of structural and descriptive metadata, at varying scales and to multiple purposes, in performing algorithmic analysis of media.

However, nowhere in film studies has descriptive metadata been more important than the practice of filmography, particularly

as it relates to the archiving and study of silent cinema. David Pierce's finding that some 75% of all American silent features are either completely lost or incomplete, in addition to being just one of the more recent pieces of scholarship made possible by large-scale metadata analysis, helps to explain metadata's particular importance in that subfield.[5] Because so much of American silent cinema is lost, filmographies and the metadata they contain—culled from archival documents, secondary sources, and the trade press—might represent the sum of all surviving knowledge about some films. Indeed, Pierce's figure does not even account for the roughly 30,000 short films of three or fewer reels produced between 1908 and 1920, which constituted the standard format of commercial filmmaking before around 1915.[6] For historians interested in such films, or the wider industry that produced them, filmographic metadata serves as a kind of cultural fossil record, representing thousands of extinct species that we can never see in the flesh.

This essay details the transformation and eventual digitization of one of the most important sets of filmographic metadata for silent cinema history: Einar Lauritzen and Gunnar Lundquist's *American Film-Index*, first published in 1976 and later indexed by Paul Spehr for *American Film Personnel and Company Credits, 1908–1920*.[7] I have transformed and uploaded Spehr's data for online open access and revision at Early Cinema History Online (ECHO). Through this case study, this essay seeks to answer a number of specific questions on the general theme of credits metadata's applicability to historical method. How does the process of digitization change the kinds of historical questions film historians can ask of their metadata? Are large-scale digital filmographic databases simply more powerful resource discovery tools, or can they themselves become the object of historical inquiry? What are the limitations of such databases, and how might we overcome or mitigate them? Ultimately, while the answers to each of these questions depend on the set of metadata in question, I would

argue that cinema historians—and particularly scholars of early cinema—can and should excavate productive knowledge from scaled digital metadata.

MAKING CREDITS ANALYZABLE: ON THE DIGITAL TRANSFORMATION OF METADATA IN MEDIA HISTORY

One of the great advantages media studies enjoys with regard to digital methods is that in certain respects, much of the basic work of *collecting* historical metadata has been done. The media industries themselves have long collected scaled datasets recording their collective activities, commodities, and personnel. Such sets were released in a variety of physical forms, from broadcasting schedules and programming guides to yearbooks, annuals, and catalogs, but they served a common goal within these industries: marketing products to consumers and audiences. Much media history is recorded in these industry-produced documents; however, because of their role in selling cultural products, they were often not consciously organized in a manner conducive to historical research and reference. Where media history's metadata is scattered across multiple documents in a haphazard way—as is usually the case in early cinema filmography—historians have stepped in to organize, compile, and index it.

Consider the two volumes of the *American Film-Index*, compiled by Einar Lauritzen and Gunnar Lundquist. Paul Spehr described the origins of the *Film-Index*:

> Lauritzen and his friend Lundquist were fans of American silent film and were unhappy that so little specific information existed about films from the silent era. They used Lauritzen's personal library to compile production information (titles, companies, dates, and existing credits) for American films made and distributed during those years. 1908–1920 was chosen because Lauritzen [...] had a complete set of *Moving Picture World* minus vol. 1 (1907). When Vol. 2 [of the *Film-Index*] was

> published, Lauritzen showed it to his friend Gösta Werner, who commented, "That's very good, Einar, but you call it an index—where's the index?" Lauritzen was in his 80s and unwilling to take on indexing. He asked me to find someone who could do it. Lacking any volunteers, I took it on. I finished in 1995, and it was published by McFarland & Co. as *American Film Personnel and Company Credits, 1908–1920*.[8]

The process Spehr describes here, which played out over the course of twenty years with the aid of collaborators Larry Karr and Susan Dalton, may seem like a simple record of the *Film-Index* data's provenance. However, a deeper examination of Spehr's indexing process places into relief some of digitization's practical effects on the usability of metadata, effects that have important methodological ramifications for the practice of media history in the digital age.

Even though Spehr's ultimate goal was the static text of a physical index, he needed to digitize the *Film-Index*'s metadata for each individual film title before he could organize and construct that index using dedicated software.[9] This involved the manual entry of credits data into a DOS-based database management system, dBase. This manual indexing was only the first of the digital transformations of Lauritzen and Lundquist's metadata, but it was by far the most important, since it required intensive and long-term human intervention. This first transformation did not simply put the physical text of the *Film-Index* into digital form; rather, it constructed a clearly defined set of digital relationships between film titles and their associated metadata, including personnel and company credits, dates, and references to pages of *Moving Picture World*. The process of that transformation served a very specific end: organizing what was effectively a giant alphabetical list of film titles with credits (the *Film-Index*) into a true index (*American Film Personnel and Company Credits*).

For scholars who had previously been unable to use the *Film-Index* to systematically and economically generate credits metadata for the specific early cinema personnel or companies they were researching, Spehr's text made possible a new set of historical questions and clarified a new set of historical relationships. Even more fundamentally, however, his work made the *Film-Index*'s valuable metadata machine-readable *as* metadata, not simply as undefined text. The defining of hundreds of thousands of individual relationships between digital entities could not have been accomplished through optical character recognition or other automation. It required years of human labor, not only on the part of Lauritzen and Lundquist to produce in the first place, but on the part of Spehr and his collaborators to properly index. Media historians who work with metadata would thus do well to remember that at its root, metadata represents a series of human-defined relationships—and that as a result, it is subject to human biases, misunderstandings, and oversights.

Media historians should also note the extensibility of Spehr's work. The dBase database was easily indexable according to specific research questions that go beyond personnel and companies (reel and release data, for instance), and Spehr has always been generous in sharing and reformulating the data for researchers.[10] But the transformation of the *Film-Index* data into digital form made it extensible to a broader set of research questions than even Spehr had anticipated. This extensibility has to do with the fact that dBase outputs its records in a consistent text format, as demonstrated by this record for Reginald Barker's *A Tragedy of the Orient* (1914), starring Sessue Hayakawa and Tsuru Aoki:

```
Record: 20229

ID                    00031834
TITLE                 Tragedy of the Orient, A
COMP                  Broncho
```

```
DATE                    1914/06/13
NAME                    Broncho (com)
NAME                    Barker, Reginald (dir)
NAME                    Chatterton, Lucille (aut)
NAME                    Borzage, Frank (cas)
NAME                    Hayakawa, Sessue (cas)
NAME                    Osborne, George (cas)
NAME                    Curse of Caste, The (ati)
NAME                    Aoki, Tsuru (cas)
DTSC                    m
VOL                     1
PG                      622
```

A record number separates each individual title record, and each data field is represented on a separate line. Descriptive metadata about each field's entry was contained in parentheses, with consistent abbreviated descriptions for each relationship (*dir* for director, *cas* for cast member, and so forth).

When Spehr kindly shared this data with Project Arclight and Media Ecology Project in 2014, all of the *Film-Index* records were contained within a single text file. The consistent format of each record allowed me to write a Perl script to parse all of the 35,000 title records in Spehr's database and convert them from raw text (.txt) to extensible markup language (XML), a standard formatting language for metadata (the current public form of this data is available at http://echo.commarts.wisc.edu/). Thus, the added value of Spehr and others' labor of digitization, beyond the end product of a true index of Laurizen and Lundquist's work, was a consistently formatted dataset (in the form of dBase) that allowed later researchers to repurpose and extend its contents. This should serve as a model for current and future media historians producing scaled metadata. Scholars must ensure that datasets can be exported in a consistent format, regardless of the form the data took when it was originally entered (or its form when ulti-

mately published). Even though dBase is largely obsolete today, and we had no access to the software used to originally input and index the *Film-Index* data, we were able to extend its usability with a minimal amount of human labor simply by parsing the text. That is, metadata's extensibility might be enhanced by digital tools or reformatting, but it is also inherent to the data itself, and scholars who preserve that quality for long-term reusability are doing a great service to media history.

Once in XML, the *Film-Index* data became convertible to any number of different formats, and indexable according to any set of multiple metadata fields (beyond title, company, or personnel) using customized queries. Film historians using this dataset can now organize film titles by release date *and* company, return a list of all Broncho films directed by Reginald Barker in 1914, or generate a list of films credited to the scenarist Lucille Chatterton. This flexibility opens the data to researchers working on very precise questions. Indeed, through every stage of its digital transformation, the field of inquiry made possible by the *Film-Index* dataset has widened; from its beginnings as a reference resource ("Which company produced this film?"), it has become a historical text open to interpretation and analysis ("What percentage of credited scenarists in the *Film-Index* are women?"). What methodological implications does this widening have?

For one, it requires that media historians think of metadata simultaneously as a tool for resource location and as itself an object of study—but, in both instances, also as information with a provenance, lacunae, and interpretive challenges. Despite its seeming comprehensiveness, the *Film-Index* was based on a single American trade periodical, *Moving Picture World*, and as a result it contains almost no information on actualities or animation. As is bound to happen in a filmography of so many films that was worked by so many hands, it also contains mistakes and omissions, some of which are easily corrected (typographical errors

and transpositions) and others that are not (misattribution). This knowledge is crucial to our critical understanding of the limitations of the dataset, and consequently the limitations of the questions we can ask it. By way of a practical example, we might use the ECHO data as a reference source to generate a list of named entities according to particular criteria—perhaps of all films released by the Fox Film Corporation in 1918. Or, we might look across the years 1914–20 to see how many films Fox released in each year, treating the metadata as a corpus for analysis. In both cases, we would need to recognize that our results would not include the popular *Mutt and Jeff* series of animated shorts, skewing our results. Media historians must thus adopt an explicit interpretive framework when working with scaled metadata, as well as having a solid understanding of the provenance of their particular set.[11]

Another methodological adjustment historians must make in using metadata as an object of analysis is to treat the queries submitted to the data with as much scrutiny and precision as they (ideally) use for the data itself. This scrutiny must be technical as well as methodological; knowing one's tools is at least as important as knowing one's methods. Depending on the researcher's mode of accessing the dataset, explicit search queries (written in SQL or xQuery, for example) might be needed. In such cases, queries should be carefully tailored to account for ambiguous entities such as name variants, alternate titles, or uncertain dates; in the *Film-Index* data, for instance, variations on D. W. Griffith's name are common. Or, if a researcher is employing a more user-friendly keyword search interface, they should consult available documentation to determine how the particular interface handles queries, string literals, Boolean operators (OR, AND), and wildcard or other special characters (*).

Finally, media historians should embrace a collaborative and revisionist approach to scaled datasets. The data that now inhabits

ECHO is the result of painstaking work from many scholars over the course of some thirty years, and it has taken its current form as a result of the willingness of its curators to share it with others for further improvement and refinement. Paul Spehr, in particular, deserves much of the credit for this collaborative process, as he has shared both the full set and distillations of the *Film-Index* with many scholars (including me) over the years. Encouraging this kind of collaboration requires a new approach to the scholarly valuation of data and text collection and curation in the humanities. Digitizing, processing, and indexing scaled corpora of information, contrary to popular belief, is not merely technical work—it is a scholarly practice, as Jerome McGann and others have argued.[12] This is especially true with regard to metadata, because metadata is directly subject to the research interests of scholars. Data about data is only recorded in the first place because someone is interested in some aspect of the original object; its collection is an active process, not a simple act of transcription. Only if we as historians acknowledge that this activity is valuable—and never as "finished" as published books and articles often seem to indicate—can we take full advantage of our metadata. From data collection to publication and back again, collaboration and revision are fundamental requirements of digital scholarship.

METADATA ANALYSIS IN ACTION: ALGORITHMIC FINDINGS FROM ECHO

ECHO's early cinema credits database was built with the principles of access, collaboration, and revision in mind. Users are able to create an account on the site, suggest and input revisions and additions to the *Film-Index* dataset, and openly access and download any portion of its data via an application programming interface (API).[13] Before ECHO's online interface existed, however, the data existed in digital form only on a handful of local machines—including my own personal computer, which I used to format and upload the dataset. As I worked with the data, I was struck by its scale and relative comprehensiveness, and decided to investigate

what riches it might hold for my own research interests in the early years of the Hollywood studio system. As it turns out, the *Film-Index* dataset offers a number of interesting algorithmic findings on company production throughput, distribution strategies, and writing credits for women in American cinema during the 1910s.

The process used to derive these findings is worth describing briefly. Using xQuery, an XML-based querying language, I was able to return lists of metadata from the *Film-Index* set according to various indexing criteria. For example, in order to get a sense for the production throughput of various companies in the 1910s, I submitted a query for all company names in the set, by year. This returned thirteen lists (one for each year) of thousands of instances of the <company> XML element, with each instance representing one title produced by that company.[14] I then processed each list in the command line with sort and uniq –c UNIX functions, in order to count the number of instances, and exported that data into a comma separated values (CSV) table. The result was effectively a manifest of company production by year for the period 1908–20 as recorded in the *Film-Index* and by Spehr, and it could be visualized in standard spreadsheet software. Below is a chart of that data, tracking the number of titles released by year for the top twenty-five most productive companies during the period 1908–20 (figure 1).

To be clear, this chart should not be taken as a foolproof, definitive listing of company output in the teens, but as a representation of the *Film-Index* dataset. There are numerous ambiguities and likely a few errors in it. For example, the single film attributed to Fox in 1914, *Life's Shop Window*, was technically produced by the Box Office Attractions Company, but at some point in the life of the set a filmographer decided to attribute it to that company's corporate successor, the Fox Film Corporation. In fact, Box Office Attractions is mentioned nowhere in that record. The set is also

Company	1908	1909	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	Grand Total
Vitagraph	75	138	156	209	302	366	337	351	249	199	182	139	137	2840
Lubin	54	108	109	167	270	342	366	336	128					1880
Essanay	38	67	114	166	226	274	299	326	150	127	20	1		1808
Kalem	41	48	96	132	190	250	256	314	273	107	2			1709
Selig	32	72	102	156	230	293	284	243	114	94	6	2	16	1644
Edison	52	121	115	194	215	265	280	187	22	48	3			1502
Biograph	60	146	118	132	137	203	244	239	35	1				1315
Universal						2	37	81	85	106	140	242	308	1001
American			11	122	131	150	114	171	116	53	17	2	3	890
IMP		9	100	115	175	151	102	93	99	45				889
Nestor	1		48	78	161	161	99	102	85	57	50	13		855
Pathé			21	45	155	51	16	2	46	89	98	157	174	854
Thanhouser			76	112	134	145	124	135	48	9				783
Reliance			9	107	106	132	167	168	3					692
Paramount									15	175	167	186	139	682
Triangle								12	195	248	139	31		625
Bison		35	101	103	79	91	51	53	51	39				603
Powers		9	100	119	107	101	77	55	25	1				594
Rex				43	98	102	106	84	78	30				541
Fox							1	33	53	102	92	85	124	490
Majestic				6	112	131	124	98						471
Keystone					29	135	151	108	8	5	2			438
Metro							1	39	98	124	78	46	45	431
Victor					26	59	97	79	69	83				413
World							20	86	69	57	54	56		342
TOTAL	353	753	1276	2006	2883	3404	3353	3395	2114	1799	1050	960	945	24,292

Figure 1. Number of titles, 1908–20, recorded in *Film-Index* data of top twenty-five most productive companies. Companies sharing a major distributor are highlighted by color: green for the General Film Company, blue for Universal, orange for Mutual, red for Triangle.

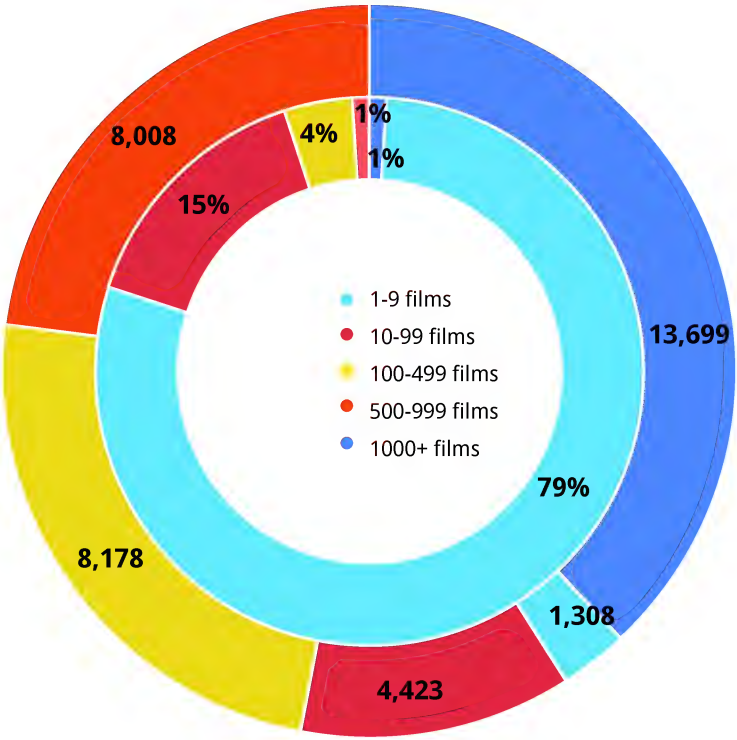


Figure 2. Production output, 1908–20, by percentage of companies (inner ring) and total number of films (outer ring).

loose in its definition of a title's associated "company," with no distinction being made between a film's production company and its distributor. Many of the films credited to Triangle—which was primarily a distributor—were actually produced by Majestic, the New York Motion Picture Company, or Keystone. This points to the vast potential for expansion, revision, and improvement in the *Film-Index* set.

However, the admittedly imperfect quantitative data in the set does illuminate interesting historical patterns when broken down. A clear shift in production took place in the period 1915–17, as

feature-oriented companies like Paramount, Fox, and Metro began making more films and many of the shorts companies that had dominated production in the early teens either shifted toward features or ceased to exist. The sheer prolificacy of the major producers during the one-reel period (roughly 1907–15) is also highlighted, if only by the fact that the top eight most productive companies—seven of which were members of the Motion Picture Patents Company—accounted for almost 40% of all the films in the set. The fact that Vitagraph tops this list points to the extent to which it was able to successfully survive the breakup of the MPPC and navigate the transition to features. This data can thus prove useful not only as an illustration of macro-level industry shifts, but also as a starting point for scaled discussions of individual companies' production.

At the same time, a more holistic look at the data reveals the "long tail" of production during this period and troubles some received historical notions about film distribution in the 1910s. Two-thirds of the individual companies recorded in the *Film-Index* metadataset produced only one or two films between 1908 and 1920, and a solid majority (more than 70%) of the 700 or so films from such companies were released after 1914. To be sure, this long tail is also very thin; the represented companies account for only about 2% of all 35,000 titles in the set. Yet a closer look at them reveals a diversity of production types. The films they produced were not all obscure one-offs, and they ran the gamut from the period's huge special features (such as 1915's *The Birth of a Nation* and 1916's *Civilization*), to independent features from well-known stars (Hobart Henley's 1918 *Parentage*, produced by Frank J. Seng), to films we know virtually nothing about (1916's *Carma*, directed by John Harvey and starring Sylva Carmen, from Florida Productions).

As named entities, the diversity of these companies shows one of the perils of that recurring theme of digital humanities, namely

distant reading, as a sea of quantitative scale tends to drown out qualitatively crucial context.[15] However, the sheer number of these companies prompts us to examine them closely, both as a totality and as definable groups, and to search for any shared historical characteristics. And at least one feature links a great many of them: distribution via alternative means, whether through state rights, roadshowing, or as a special through one of the major companies. That a majority of these films were produced in 1915 or later suggests that the importance of alternative distribution in the aftermath of the wider industry's transition to features may have been underestimated. Current histories tend to frame the late teens as a period of industry consolidation and vertical integration; this data forces us to consider the extent to which marginal forms of distribution continued to be important for certain producers and under certain circumstances.[16]

Outside of the company data, the *Film-Index* has valuable evidence to contribute to contemporary scholarship on women screenwriters during this period. Scholars like Shelley Stamp, Amelie Hastie, Mark Garrett Cooper, and Jane Gaines have documented the important creative roles that women played in early Hollywood, illuminating the forgotten contributions of scenarists in particular.[17] As Stamp points out, women authored a significant proportion of silent screenplays.[18] The *Film-Index* dataset helps put this in a quantitative context. While the gender of credited entities was never directly coded into the set—another arena in which it might be improved—a basic culling of the 10,075 individual scenario credits was possible through the same xQuery and Unix-based method employed to produce a manifest of company production. Using this method, I output a text list of all 1,981 individual screenwriters along with the quantity of their credits. By going through this list with the aid of the Women Film Pioneers website and additional name coding, the absolute minimum percentage of films written by women could be calculated. Out of the individual scenario credits in the set, 1,576 (16%) are unambiguously for

women, representing 332 named screenwriters (a roughly equal proportion, 16%, of all screenwriters). This percentage represents an absolute minimum, and does not factor in ambiguous names, of which there are many, and of which the majority may very well be for women writing under intentionally ambiguous names. Furthermore, the overwhelming lack of authorship documentation for shorts, where women's scenario-writing labor would have gone largely unrecorded, additionally suggests a significantly higher percentage than 16%. This would seem to support Anthony Slide's contention that women wrote 20–25% of silent productions in the United States.[19]

Examples such as this show the power of quantitative data to confirm or nuance existing scholarship in media history, and a forthcoming study using the credited women mentioned in the ECHO database finds that screenwriters like Anita Loos and June Mathis were particularly prominent in the Hollywood trade press in terms of the amount of coverage they received.[20] Of course, data of this kind can only serve as a starting point for further histories of women's contributions to early cinema and early Hollywood. It tells us precious little about questions of genre, characterization, performance, or any other of the hundreds of qualitative elements that would be a necessary part of such histories. But one of the most exciting prospects of the *Film-Index* dataset as it has been transformed for ECHO is that such metadata could conceivably be coded into the records of individual films for further analysis and interpretation. The extensibility ECHO enables via user corrections and contributions, along with its open-access model, makes such analysis and study simultaneously possible, public, and shareable. While the initial quantitative findings offered here are relatively modest, and largely support established scholarship on early cinema and Hollywood history, the methods used to derive these findings point to the possibility for more extensive revisionist scholarship on the American film industry as a scaled entity.

CONCLUSION

While previous versions of the *Film-Index* data have been used by other scholars, ECHO represents the first comprehensive effort to mount and work with that data in an entirely digital context. It is also the first time that the data has been compiled for long-term use and collaborative improvement by the community of early cinema scholars and amateur enthusiasts. In this article, I have used the example of ECHO as a model for the kind of metadata compilation and analysis that digital formats and methodologies make possible. As we have seen, metadata analysis is a fruitful approach to digitized corpora—an approach that requires more theorization, to be sure, but also one that calls for technical implementation and active research programs. If media scholars are prepared to value the labor of digitization in a collaborative, revisionist way, they may discover that inside the cultural fossil record represented by more than a century of the media industries' collective metadata, there lies a trove of undiscovered species.

ENDNOTES

1 It is worth pointing out that while librarians and information scientists might have stricter definitions of "metadata" than simply "data about data," the use of the term tends to be very functional and contextual in the digital humanities. A piece of information that constitutes metadata in one context—that is, it is information that relates to the object or evidence being studied—might very well be data in another, becoming the object of study. Metadata analysis is thus fundamentally related to this process of transformation of metadata into data, and often involves activities of preprocessing to make metadata researchable as data.

2 Rebecca Guenther and Jaqueline Radebaugh, *Understanding Metadata* (Bethesda: National Information Standards Organization Press, 2004), 1.

3 See Rick Altman's "Table of Musicals by Subgenre" in *The American Film Musical* (Bloomington: Indiana University Press, 1987), 371–78.

4 See, for example, Raymond Bellour's canonical segmentation of Hitchcock's *North by Northwest* (1959), which breaks down the entire film

according to the various means of locomotion employed by protagonist Roger Thornhill. *The Analysis of Film* (Bloomington: Indiana University Press, 2000), 175–78.

5 David Pierce, *The Survival of American Silent Feature Films: 1912–1929* (Washington, DC: The Council on Library and Information Resources and the Library of Congress, September 2013), 21.

6 See Eileen Bowser, *The Transformation of Cinema, 1907–1915* (New York: Scribner's, 1990).

7 Stockholm: Film-Index, 1976; Jefferson: McFarland, 1996.

8 Jefferson: McFarland, 1996. Personal email correspondence with Paul Spehr, June 27, 2014. Thanks to Paul Spehr and Mark Williams for sharing this correspondence with me.

9 Spehr, email correspondence, June 27, 2014. According to Spehr, the indexing program, IPS, was created by Larry Karr for the National Park Service.

10 Ben Singer, for instance, based much of his quantitative work on American feature and shorts production in the 1910s on statistical breakdowns generated by Spehr and Karr. Ben Singer, "Feature Films, Variety Programs, and the Crisis of the Small Exhibitor," in *American Cinema's Transitional Era*, ed. Charlie Keil and Shelley Stamp (Berkeley: University of California Press, 2004), 77–78, 96 (footnote 3).

11 Hoyt et al. in "Scaled Entity Search: A Method for Media Historiography and Response to Critiques of Big Humanities Data Research," *Proceedings of IEEE Big Humanities Data* (2014), https://bighumanities.files.wordpress.com/2014/10/hoyt.pdf, offer one such interpretive framework, Scaled Entity Search. Applying SES to metadata would require researchers to think of their dataset as both a set of query entities and the corpus being queried.

12 Jerome McGann, "Philology in a New Key," *Critical Inquiry* 39:2 (Winter 2013), 327–46; Eric Hoyt, Derek Long, Anthony Tran, and Kit Hughes, "Variety's Transformations: Digitizing and Analyzing the first 35 Years of the Canonical Trade Paper," *Film History* 27:4 (2016): 76–105.

13 Users of ECHO's search interface can also download their search results in XML and other formats.

14 Each title's XML record contains only one `<company>` element,

even if the company is listed as a `<namedEntity>` again in the same record. Using the `<company>` element, rather than simply all instances of a company name, prevents multiple instances from being returned for a title.

15 Lev Manovich, "How to compare one million images?" in *Understanding Digital Humanities,* edited by David M. Berry (New York: Palgrave Macmillan, 2012), 249–98.

16 See Richard Koszarski, *An Evening's Entertainment: The Age of the Silent Feature Picture, 1915–1928* (Berkeley: University of California Press, 1990), 63–94.

17 Shelley Stamp, *Lois Weber in Early Hollywood* (Berkeley: University of California Press, 2015); Amelie Hastie, *Cupboards of Curiosity: Women, Recollection, and Film History* (Durham: Duke University Press, 2007); Mark Garrett Cooper, *Universal Women: Filmmaking and Institutional Change in Early Hollywood* (Champaign: University of Illinois Press, 2011); Women Film Pioneers Project (http://wfpp.cdrs.columbia.edu/).

18 Stamp, Lois Weber, 24.

19 Anthony Slide, "Early Women Filmmakers: The Real Numbers," *Film History* 24:1 (2012): 114–21.

20 Derek Long et al., "Who's Trending in 1910s American Cinema? Exploring ECHO and MHDL at Scale with Arclight," forthcoming in *The Moving Image,* 2016.

BIBLIOGRAPHY

Altman, Rick. *The American Film Musical.* Bloomington: Indiana University Press, 1987.

Bellour, Raymond. *The Analysis of Film*. Bloomington: Indiana University Press, 2000.

Bowser, Eileen. *The Transformation of Cinema, 1907–1915*. New York: Scribner's, 1990.

Cooper, Mark Garrett. *Universal Women: Filmmaking and Institutional Change in Early Hollywood*. Champaign: University of Illinois Press, 2011.

Guenther, Rebecca and Jaqueline Radebaugh. *Understanding Metadata.* Bethesda: National Information Standards Organization Press,

2004.

Hoyt, Eric, Kit Hughes, Derek Long, Kevin Ponto, and Anthony Tran. "Scaled Entity Search: A Method for Media Historiography and Response to Critiques of Big Humanities Data Research." *Proceedings of IEEE Big Humanities Data.* 2014. https://bighumanities.files.wordpress.com/2014/10/hoyt.pdf.

Hoyt, Eric, Derek Long, Anthony Tran, and Kit Hughes. "*Variety*'s Transformations: Digitizing and Analyzing the First 35 Years of the Canonical Trade Paper." *Film History* 27, no. 4 (2016): 76–105.

Koszarski, Richard. *An Evening's Entertainment: The Age of the Silent Feature Picture, 1915–1928.* Berkeley: University of California Press, 1990.

Lauritzen, Einar and Gunnar Lundquist. *The American Film-Index.* Stockholm: Film-Index, 1976.

Long, Derek, Eric Hoyt, Kevin Ponto, Tony Tran, and Kit Hughes. "Who's Trending in 1910s American Cinema? Exploring ECHO and MHDL at Scale with Arclight." Forthcoming in *The Moving Image,* 2016.

Manovich, Lev. "How to Compare One Million Images?" In *Understanding Digital Humanities,* edited by David M. Berry, 249–98. New York: Palgrave Macmillan, 2012.

McGann, Jerome. "Philology in a New Key." *Critical Inquiry* 39, no. 2 (Winter 2013): 327–46.

Pierce, David. *The Survival of American Silent Feature Films: 1912–1929.* Washington, DC: The Council on Library and Information Resources and the Library of Congress, September 2013.

Singer, Ben. "Feature Films, Variety Programs, and the Crisis of the Small Exhibitor." In *American Cinema's Transitional Era,* edited by Charlie Keil and Shelley Stamp, 77–78, 96 (footnote 3). Berkeley: University of California Press, 2004.

Slide, Anthony. "Early Women Filmmakers: The Real Numbers." *Film History* 24, no. 1 (2012): 114–21.

Spehr, Paul. *American Film Personnel and Company Credits, 1908–1920.* Jefferson: McFarland, 1996.

Stamp, Shelley. *Lois Weber in Early Hollywood.* Berkeley: University of California Press, 2015.

Women Film Pioneers Project. http://wfpp.cdrs.columbia.edu/.

SHOW ME THE HISTORY! BIG DATA GOES TO THE MOVIES

Deb Verhoeven

> "To have lived so long with time and to find, when one thought one had *all the time in the world*, that time had deserted, disappeared."
> Janet Frame[1]

> "The apocalypse is not something which is coming. The apocalypse has arrived in major portions of the planet and it's only because we live within a bubble of incredible privilege and social insulation that we still have the luxury of anticipating the apocalypse."
> Terence McKenna[2]

ALL THE TIME

Before big data could up anchor and put to sea, before it had unfurled and stretched its sails, the wind was already turning. Hasty verdicts disdaining the utility, merit, influence, and defining features of large-scale data driven studies abruptly deflated the ascent of big data's hype curve. The preemptory backlash against big data[3] has been especially pronounced in the humanities and creative arts[4] where amplification alarmism[5] and concerns over historical ethics and methodologies[6] have prompted calls for extreme caution.

In this context there has been little concerted effort made in the humanities and creative arts to specifically assess how big data might contribute something, anything, to the way we undertake data-driven enquiry. In film studies particularly, the recent availability of very large datasets has the potential to alter the shape and scope of our studies, as well as prompt us to creatively reconsider some of the underlying precepts and practices of our research.

This chapter is based on my work within the Kinomatics Project, a multidisciplinary big data study of film exhibition at an international scale. The project proceeds from the New Cinema History's proposition that cinema is not an isolated set of practices but comprises institutional, social, and commercial networks that are interdependent and which in turn influence and shape our own approach to the field.[7] So for example, the Kinomatics Project combines its central film exhibition dataset with other data (demographic, social media, technical infrastructure, economic and financial, and climatic data, to name just a few) in order to explore the value of an expanded approach to cinema data, rather than simply focusing on the idea of one big dataset per se. In this sense, the Kinomatics Project demonstrates that it's not how big your data is, it's what you do with it that counts.

Indeed, focusing only on the size of big data can result in missing its most significant features. Given the unprecedented and rapid expansion of data production, this week's big data is almost certainly going to be next week's iota. And this variability applies across different disciplines as well. The Kinomatics Project's big cinema data, for example, is tiny compared to the data used by astronomers, but it stretches capacity within the field of cinema studies. In this sense, big data might be understood as a collection of data that, in any given context, is so large that it is ungraspable and incomputable using conventional approaches to analysis. Big data is data that in some way defies our comprehension and exceeds our capacity to handle it. Instead, new adaptive computational techniques that are designed to operate within indeterminate environments are required. This aspect of big data has epistemic implications (pushing at the edges of what can be known and how we can know it) as well as ontological ones (in its reliance on machine-based analytics rather than human-centered methods).

So given its ontological and epistemological dimensions, it is not

surprising that much of the anxiety around big data is tied up in apocalyptic or originary narratives. The impact of big data in this sense is that it challenges us, through its massive presence, to imagine what in the end we are dealing with and how we are, ourselves, redefined by it. This is big data as both self-effacing and conceiving—an impossible moment of perceptual mastery, production, and knowledge in which multiple contingencies of time are condensed, and time and endings/beginnings are conflated into a totalizing coincidence. This characterization of big data is at the heart of popular arguments that suggest it will lead to the end of theory such that representation and deliberation would be made obsolete by a sheer mass of information.[8] In the discourse of data hyperabundance, big data gestures at a sort of vanishing point of history, an amniotic abstraction where differences, disparities, and divergences (the conditions of classification) disappear. In this view, big data and its information overload threaten to produce, through sheer incalculable scale, a type of invisibility or indistinguishability, an undifferentiated (zero and) Oneness from which identity, our 'selves,' might be mercurially discerned but into which they might equally disappear.

Yet neither the apocalyptic (big data as the end of the world as we know it) nor the originary (big data as a primal scene) are especially helpful frameworks for getting to how we, as *researchers* and even more specifically as film and media historians, can conceptually and practically engage with large-scale databases as part of our research repertoire. How might we better understand and perhaps intervene in the development of emerging data-driven practices? How might we aim for working with digital archives and databases as a form of historical thinking, to reflect, for example, on how the technologies we engage in might also be attributed temporalities, that they do not simply and instrumentally and chronologically follow a preexisting claim for the truth? We might instead take this opportunity to better consider how different computational technologies participate in and respond to chang-

ing definitions of time and history. In amassing and archiving vast amounts of commercial cinema exhibition data that would otherwise be disposed of, scholarly projects such as Kinomatics create new forms of research repository that invite new uses, practices, and questions. These include examining the kinds of change and continuity that are already inscribed temporally within big data and which might contribute to a revised understanding of what we mean when we talk about film history.

And yet, because the Kinomatics Project is perceived by many to be a study of contemporary cinema, it is frequently isolated from the New Cinema History which forms its intellectual framework. More often than not, the work we are undertaking is characterized as being not "prior" enough, our data not sufficiently dated to contribute to matters of history; as if time is a stream that flows forward in one direction, coursing from the headwaters of the past through the present to estuaries of the possible; as if there is some identifiable point in time that segregates the past from the present; as if time is external, an abstract measure that can be applied to our studies and which lays down the syntactic rules for determining scholarly disciplines by progressions of tense.

However, if we understand both our information systems and our disciplines as inherently theoretical and temporary formations/formulations, then we can also consider what theoretical and historical questions they themselves recommend and advance. And then in turn, how our own understandings of (new) cinema history might contribute to a practical reconsideration of emerging digital research techniques. All disciplines are temporal gatherings, bearing ideas about the past and the present, of what was and what is (and usually an implicit sense of what should be): a notion of time, a theory of history. With this in mind we can consider how working with the Kinomatics data practically and theoretically alters the "new," the "cinema," and the "history" in what new cinema historians do.

By its very nature working with large datasets challenges the tendency to taper history to a specific temporal horizon or to a belief in chronological succession. By insisting we analyze *at scale*, rather than using proxy datasets as a metonymy for interpretative generalizations, big cinema data brings to light the ways in which multiplicities and complexities of time actively produce film history; in which, for example, the nature of film and the film industry contribute to the production of time; and the ways in which data and the databases that accommodate them also lend themselves to the production of differing dimensions of temporality. In this essay I want to show how film history can be seen emerging from a set of uneven, variable temporalizing processes rather than as a set of sequential points known distinctly as past, present, and future.

This chapter will explore, in the context of the Kinomatics dataset, both how cinema researchers can work with historical data and how cinema researchers can work historically with data. I want to consider how the experience of using big data opens up more nuanced ways of thinking temporally and historically around our digital archives and databases. I want to ask how data-based research collections like Kinomatics might inspire researchers to reflect on the nature of history and how we might deal differently with passing media, passing computational technologies, and also passing ideas about pastness itself (what is it, when is it, who or what gets to exist in it, and who decides?).[9]

IN THE WORLD

To date, digital cinema exhibition and distribution history has been undertaken through a series of initiatives produced "from below."[10] Without exception the existing datasets that form the empirical basis for digital cinema research have occurred at the national or subnational level. Cinema datasets have been generated for scholarly research projects focused on (and not limited to): London,[11] the Netherlands,[12] Ghent,[13] Antwerp,[14] Australia,[15]

Scotland,[16] Italy,[17] and North Carolina.[18] Each of these datasets was developed independently to solve specific research problems and they are not technically or semantically compatible. The prospect of interoperating these data collections remains a tantalizing but near impossible challenge with few options for resourcing an undertaking of this magnitude.

While the proliferation of these digital case studies has produced a great deal of methodological innovation in cinema studies, this disjointed approach has also resulted in a significant deficit in our understanding of the international nature of the cinema. These distributed research collections are not yet capable of addressing the global, elastic, and networked nature of the contemporary international film industry that is itself currently producing and exploiting huge quantities and varieties of data. Companies such as Rentrak and Netflix, for example, are using newly available big data (describing purchasing behaviors, preferences, and social media sentiment) to drive business decisions including production investment and the customization of promotional materials to the level of individual consumers. For the first series of the US television series *House of Cards* (2013–), Netflix created ten different trailers that were circulated according to the specific viewing profiles of subscribers developed through an analysis of consumer preferences. Netflix also analyzes large-scale transactional data to improve playback quality (and understand how changes in the quality of viewing experience affect user behavior) and identify poorly translated subtitles.

The primary source of data for the Kinomatics Project derives from our global showtime database. Data arrives on a weekly basis from a third-party commercial data provider. This data records all screenings of all films for all cinema venues in forty-eight countries around the world. We collect data for formal theatrical distribution only (not, for example, community screenings or viewings in other media). Previously this information would

have been discarded as noise both industrially and within cinema studies. For film historians wanting to examine pre-digital film exhibition and distribution, extant records such as theater log books are highly coveted for their rarity. Instead researchers must typically reconstruct cinema programs from newspaper advertising and other ephemera. Our collection of monumentally detailed screening data is unique in film research to this point, and the Kinomatics showtime database is the only repository of this data in the world. Nevertheless, its accessibility to researchers outside the Kinomatics team is restricted by our legal contract with the commercial data provider. The dataset includes data about:

* Venues: name, addresses, geographic coordinates, number of screens, sound technology, etc.
* Movies: title, main actors, genre, running time, director, writer, producer, etc.
* Showtimes: film, venue, date, time, whether it was part of a film festival

The time period for the Kinomatics showtime database is December 1, 2012 until May 31, 2015. During this thirty-month period, we collected data on just under 97,000 movies playing in over 33,000 venues with a total of 338,660,831 screenings. Although the vast majority of screenings are for first release titles, the screenings recorded in the dataset are not specifically limited to new releases. The data provider obtains information directly from cinema venues mostly through automated electronic means and also email and phone calls. Once we receive the data, it is stored on a Linux server at Deakin University and then organized into a data model with a consistent format and hosted in a relational database (MySQL 5.1.67). (See figure 1 for the database schema).

Although very large, the Kinomatics data is limited in its cleanliness (in that some values are missing in some records), in its evident biases (Western commercial cinema is far better repre-

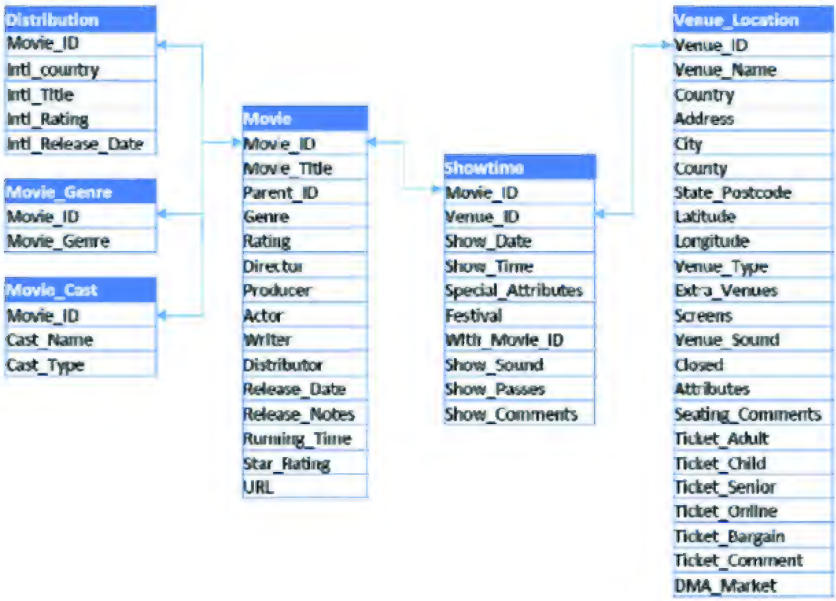


Figure 1. Conceptual Kinomatics database model.

sented), and in its consistency (standards of data collection vary for some countries). Because of the sheer size of the dataset many of these anomalies are not evident until experiments and audits are performed on the dataset (often as visualizations of the data). Indeed, working on the Kinomatics showtime dataset has entailed from the outset, of necessity and in principle, an orientation to iterative and recursive ways of working with data.

This focus on feedback is perhaps most evident in our attempts to think around the restrictions on access to the data that resulted from our commercial data contract. Two projects, the Cinema Cities project[19] and the Film Impact Rating project,[20] represent our attempts to make the Kinomatics data accessible in some way, as well as make transparent the algorithmic processes on which so much big data analysis relies. In both these projects the public are invited to engage with the dataset by expressing their own values and preferences using an online tool. In the case of Cinema Cities,

they can 'weight' their motivations for cinema venue attendance on a sliding scale, in order to produce a measure of what we call "cinemability." The combination of these weighted factors produces a ranked list of global cities that conform most closely to their preferences. In the Film Impact Rating project, site visitors can indicate their views on how any given film's success should be measured against fourteen variables including commercial attributes, critical assessments, and global venue coverage. This results in a ranked list of films based on their own weighting of success factors.[21]

These public preferences are then collected so we, the researchers, can further reflect on our own analytic decisions and choices and make adjustments. In both projects there were clear differences between the values held by the public and those proposed by the Kinomatics project team. In the case of Cinema Cities, ticket pricing was a particularly significant issue for users. For the Film Impact Rating, the public indicated that commentary such as critics' ratings and IMDb user votes, was most important to them in terms of defining a film's success. Through these participatory, feedback-focused approaches, we believe it is possible to appreciate within the global a range of diverse perspectives, inheritances, structures, and ownerships of information. In this way too, we hope that temporality emerges relationally and transparently, within a constant process of research engagement rather than the product of the conventionally imposed methods and the sequential categories of film history. This approach would certainly honor the complex temporalities of the data itself.

For the most part, Kinomatics captures data about cinema occasions that haven't yet occurred, but they may as well have occurred. The weekly arrival of data typically describes screenings from a Friday to Thursday *forthcoming*. However, because play weeks are not consistent around the world, some of the weekly data dump will describe information about screenings held on

days in a previous week. The Kinomatics database counts four types of play weeks: Wednesday–Tuesday, Thursday–Wednesday, Friday–Thursday, and Saturday–Friday (fig. 2). Known erroneous data (for instance, projected showtimes that then did not actually occur) are overwritten and corrected as they come in—so, in some cases, there is a recognized obsolescence built in to the projection of forthcoming showtimes. Kinomatics doesn't keep a copy of these replaced records so at any given time during the collection process the database is a mixture of reported and projected showtimes. Furthermore, the data provider disposes of all data after one month. So anything older than a month exists only in our dataset and nowhere else. In this sense the Kinomatics showtime database might be also considered an archival repository.

The idea of capturing prospective cinema events would seem to fly in the face of typical historical research. Data that casts into the future certainly questions the documentary impulse and truth function of conventional history as well as the claims to legitimacy of so many historical datasets. And yet, because this is the same data would-be cinemagoers see when they Google local showtimes in search of a program to attend, we can assume there is a high level of investment in its accuracy. If the forecast showtimes were incorrect then the cinema businesses issuing them would suffer.

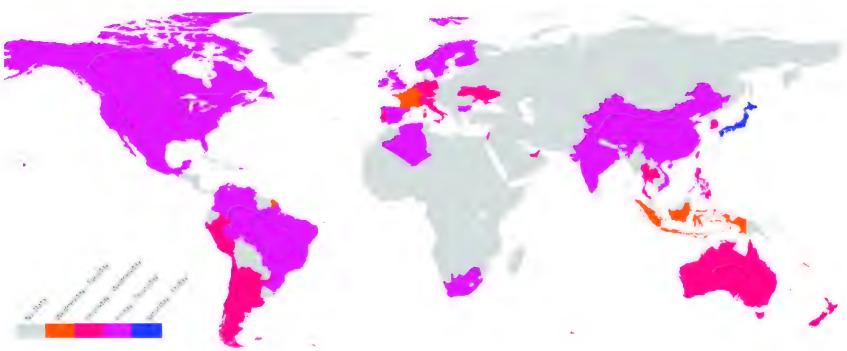


Figure 2. Differences in play weeks for countries in the Kinomatics database.

In Kinomatics, film events are "forth-comings," embodying and anticipating a particular temporality which also constitutes it as a form of archival history. Every showtime event in the Kinomatics database is made up of many information events which contain pasts, presents, and futures. Here, the lightning of computation is recreated in time's grasp as cinema data is captured at the level of intention. The projected occasions it describes provide the conditions from which the present and past are creatively assembled, in some contrast to the traditional archive's presentation of the pastness in the present. Instead, the big data of the Kinomatics showtime dataset is clairvoyant in nature, a leap of faith that reveals the politics and fragility of our capacity to know. This is the emergence of temporality in the context of constant computational processing, or, in other words, data as process, as movement. The cinema, in turn, is figured as a kind of hopeful industriousness—the result of temporal inferences and constantly evolving practices. As Georgina Born suggests, we might look within our data for "distinctive scales, speeds, rhythms, and shapes of change opened up and enacted by cultural objects and events—that through their complex interactions participate in the emergent processes we identify as history."[22] Big cinema data gives us the means to examine more open temporal systems. So, for example, we have proposed a rethinking of the annualization of analysis of the film industry.[23] And the opportunity exists to think beyond geopolitically produced temporalities (holiday seasons and so on) to accommodate other forms of temporal organization in the cinema.

There are other aspects to computational time that fall outside the scope of this brief discussion but can be at least acknowledged. Temporality is both an aspect of the various realities that databases attempt to model and it is also a form of measurement that shapes the data they keep. Many different types of time are captured by databases, and different taxonomies of computational

time (valid-time, transaction-time, user-defined time) will produce different types of demands on database design.[24] But even their measurement of time is subject to changeability. Computers and the databases they serve are not somehow above or beyond time. The past is not an outside to be captured and organized by the database or by the researcher. For Timothy Barker, working with databases (small or large) involves engaging in a process that "not only changes the information that it archives but is also generative of a particular type of presentness in which the information is accessed. This is a process that brings together pastness and presentness; a process that does not sit outside or beyond everyday life, but rather a system that is involved in a process *with* everyday life; a system that is necessarily temporal."[25] The way in which time is shaped in a digital repository is dependent on the database's organizing structures and the computer's system capabilities rather than the position of events in a linear or chronological sequence.

To begin thinking historically in data-driven research then, we might also consider how database design and file-system management produce temporal perspectives as well. A more historically informed approach to database-driven research might consider (but not be limited to) making provision for:

* **File version management and tracking** that allows researchers to see how files and directory structures have changed and evolved over time. An elaboration of this is to design for Point-in-Time Views of the file system that enable users to "turn back the clock" and see all of their data exactly as it existed at any past point in time. In an ideal world this would also entail full file system audit trails (with SQL-based reporting) that can show every change, deletion, and even access of every file in the system by every user in support of tracking activities. To accommodate historical and as-at reporting, researchers need to design their databases

with a big-data mentality, with an eye to scale and elapsed time in the construction of result sets.
* **Code versions** so that the historical development of a database at the code level is preserved on an open-access platform such as GitHub.
* **Graceful degradation** in which the web interface and functionality of the database is designed in such a way that it can continue to operate and is legible even when viewed with less-than-optimal software.[26]

CONCLUSION

Cinema archives in the form of databases present history as a complex constellation of narratives that can be searched and browsed and from which temporality emerges.

By enabling us to analyze the film industry at scale, big cultural data collections like Kinomatics bring into view different temporal dimensions, uncertainties, and contingencies. Without a doubt, my own understanding of and ability to evaluate and theorize the temporal processes of film exhibition and distribution has been challenged and changed by Kinomatics' vast network of interconnected events formed from multitemporal information.

In the context of 'big' data then, we might consider the ways that time both exceeds us and yet is not external to our historical enquiries or our selves. Working "historically" with big data should mean that our digital research efforts are as embedded, relational, and enacted as our data itself. Our ethics, methods, and theories of history should be transparent in our tools and in the way we account for their temporalities. And we must also account for the temporalities of the researcher herself, how our own personal perspectives, positions, and productions are shaped by time's grip, the way the rhythms of academic life are woven into our work, the sheer duration required to labor over large data, for example, or how patience, impulsiveness, urgency, exasperation, and for-

bearance all play a role in bringing our research to a terminus, however transitory.

The point, as a new 'big' cinema historian however, is not to bind time to our experience of it, but rather to acknowledge the ways in which our digital technologies can also iteratively open up our thinking, expectations, and encounters with time. We must endeavor to understand how we, as cinema researchers, how the technologies we work with, the film industries we study, all distribute time differently. But the work of the new 'big' cinema historian is not simply to accrue and authenticate diversities of time across the global, cultural, social, organizational, and biographical dimensions of our studies. The New Cinema History itself needs to recognize the coexistence of multilateral temporalities that are scaled unevenly between expansion (being with time) *and* contraction (being without time).

By working with large-scale digital archives like Kinomatics, we can recognize and critically reflect on how both our conventional disciplinary and technical standards have acted temporally to regulate and chronologically direct our data toward the idea of a more capacious (better informed) future. In developing new digital formats for historical research that are specifically designed to realize the temporal potential and creativity of data relations, we can now contemplate the scalability of time itself and not just our data. We can wonder what it is to simultaneously hold and be held, more or less by time; and likewise, by more or less time.

* I would like to acknowledge the efforts of the Kinomatics research team and the assistance of Kevin Whitesides, Mark Pesce, and James Verhoeven in the preparation of this chapter.

ENDNOTES

1 Janet Frame, *In the Memorial Room* (Melbourne: text publishing, 2013), 108 (emphasis in original).

2 “Imagination in the Light of Nature (Terence McKenna),” YouTube video, 1:37:13, from a presentation at the Earth Trust Benefit, October 2nd, 1992, posted by “Deus Ex Mckenna ~ Terence Mckenna Archive,” July 9, 2011, https://youtu.be/IERPnScenBM?t=8m14s.

3 “The *Economist* Explains: The Backlash against Big Data,” *Economist*, last modified April 20, 2014, accessed March 12, 2016, http://www.economist.com/blogs/economist-explains/2014/04/economist-explains-10.

4 Stephen Marche, “Literature Is Not Data: Against Digital Humanities,” *Los Angeles Review of Books*, last modified October 28, 2012, accessed March 12, 2016, https://lareviewofbooks.org/essay/literature-is-not-data-against-digital-humanities.

5 See danah boyd and Kate Crawford, “Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon,” *Information, Communication & Society* 15, no. 5 (2012): 662–79; and Anne Burdick et al., *Digital_Humanities* (Cambridge: MIT Press, 2012).

6 See Tim Hitchcock, “Academic history writing and the headache of ‘Big Data’,” *Historyonics,* last modified January 30, 2012, accessed March 12, 2016, http://historyonics.blogspot.com/2012/01/academic-history-writing-and-headache.html; and Frederick W. Gibbs and Trevor J. Owens, “Hermenuetics of Data and Historical Writing,” in *Writing History in the Digital Age,* ed. Jack Dougherty and Kristen Nawrotzki (Ann Arbor: University of Michigan Press/Digital Culture Books (Online), 2013), http://writinghistory.trincoll.edu/data/gibbs-owens-2012-spring/.

7 Richard Maltby, “New Cinema Histories,” in *Explorations in New Cinema History: Approaches and Case Studies*, ed. Richard Maltby, Daniel Biltereyst, Philippe Meers (Oxford: Wiley-Blackwell, 2011), 3–40; Deb Verhoeven, “New Cinema History and the Computational Turn,” in *WCCA 2012: “Beyond Art, Beyond Humanities, Beyond Technology: A New Creativity,”* ed. Claudio da Rocha Brito, et al. (Guimarães: COPEC-Science and Education Research Council, 2012); and Richard Maltby, Dylan Walker, and Mike Walsh, “Digital Methods in New Cinema History” in *Advancing Digital Humanities: Research, Methods, Theories*, ed. Paul Arthur and Katherine Bode (Hampshire: Palgrave Macmillan, 2014), 95.

8 Chris Anderson, “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete,” *Wired*, last modified June 23, 2008, accessed

March 12, 2016, http://www.wired.com/2008/06/pb-theory/.

9 Questions inspired by Carolyn Dinshaw's contemplation of the present in Dinshaw, *How Soon Is Now? Medieval Texts, Amateur Readers, and the Queerness of Time* (Durham: Duke University Press, 2012), 7.

10 Maltby, "New Cinema Histories."

11 Ian Christie, *The London Project*, accessed March 12, 2016, http://londonfilm.bbk.ac.uk/.

12 Karel Dibbets, *Film in The Netherlands from 1896*, accessed March 12, 2016, http://www.cinemacontext.nl/.

13 Daniel Biltereyst and Lies Van de Vijver, *Gent Kinemastad*, accessed March 12, 2016, http://www.cims.ugent.be/research/past-research-projects/gent-kinemastad.

14 Luc Pauwels, *Visual and Digital Cultures Research Centre (ViDi)*, accessed March 12, 2016, https://www.uantwerpen.be/en/rg/vidi/projects-and-publica/projects/.

15 Richard Maltby et al., *Australian Cinemas Map*, accessed March 12, 2016, http://auscinemas.flinders.edu.au/; Deb Verhoeven, Colin Arrowsmith, and Alwyn Davidson, *Cinema and Audiences Research Project*, accessed March 12, 2016, http://caarp.edu.au/.

16 Gordon Barr and Gary Painter, *Scottish Cinemas and Theatres Project: Scottish Cinemas Database*, accessed March 12, 2016, http://www.scottishcinemas.org.uk/database.html.

17 Daniela Treveri Gennari, et al., *Italian Cinema Audiences*, accessed March 12, 2016, http://italiancinemaaudiences.org/.

18 Robert C. Allen, Going to the Show: Mapping Moviegoing in North Carolina, accessed March 12, 2016, http://docsouth.unc.edu/gtts/.

19 Deb Verhoeven, Colin Arrowsmith, Bronwyn Coate, and Alwyn Davidson, *Cinema Cities Index*, accessed March 12, 2016, http://cinemacities.com/.

20 Deb Verhoeven, Colin Arrowsmith, Bronwyn Coate, and Alwyn Davidson, *Film Impact Ratings*, accessed March 12, 2016, http://www.reelmeasures.com/.

21 See Deb Verhoeven, Alwyn Davidson, and Bronwyn Coate, "Australian films at large: expanding the evidence about Australian cinema performance," *Studies in Australasian Cinema* 9, no.1 (2015): 7–20.

22 Georgina Born, "Making Time: Temporality, History and the Cultural Object," *New Literary History* 46, no. 3 (Summer 2015): 381.

23 Alwyn Davidson and Deb Verhoeven, "Bad Hobbits Die Hard: How to make a better Top 10 movie list," The Conversation, last modified December 10, 2014, accessed March 12, 2016, https://theconversation.com/bad-hobbits-die-hard-how-to-make-a-better-top-10-movie-list-35247.

24 Abdullah Uz Tansel, "Temporality in Databases," in *Advances in Information Systems,* ed. Tatyana Yakhno (Berlin: Springer-Verlag Berlin Heidelberg, 2004), 1.

25 Timothy Scott Barker, *Interfaces: Studies in Visual Culture: Time and the Digital: Connecting Technology, Aesthetics and a Process Philosophy of Time (*Lebanon: University Press of New England, 2012), 162.

26 Peter-Paul Koch, *Fluid Thinking,* accessed March 12, 2016, http://www.digital-web.com/articles/fluid_thinking/.

BIBLIOGRAPHY

Allen, Robert C. *Going to the Show: Mapping Moviegoing in North Carolina.* Accessed March 12, 2016. http://docsouth.unc.edu/gtts/.

Anderson, Chris. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," *Wired.* Last modified June 23, 2008. Accessed March 12, 2016. http://www.wired.com/2008/06/pb-theory/.

Barker, Timothy Scott. *Interfaces: Studies in Visual Culture: Time and the Digital: Connecting Technology, Aesthetics and a Process Philosophy of Time.* Lebanon: University Press of New England, 2012.

Barr, Gordon and Gary Painter. *Scottish Cinemas and Theatres Project: Scottish Cinemas Database.* Accessed March 12, 2016. http://www.scottishcinemas.org.uk/database.html.

Biltereyst, Daniel and Lies Van de Vijver. *Gent Kinemastad.* Accessed March 12, 2016. http://www.cims.ugent.be/research/past-research-projects/gent-kinemastad.

Born, Georgina. "Making Time: Temporality, History and the Cultural Object." *New Literary History* 46, no. 3 (Summer 2015): 361–86.

boyd, danah and Kate Crawford. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15, no. 5 (2012): 662–79.

Burdick, Anne, Johanna Drucker, Peter Lunenfeld, Todd Samuel Presner, and Jeffrey T. Schnapp. *Digital_Humanities*. Cambridge: MIT Press, 2012.

Christie, Ian. *The London Project*. Accessed March 12, 2016. http://london-film.bbk.ac.uk/.

Davidson, Alwyn and Deb Verhoeven. "Bad Hobbits Die Hard: How to Make a Better Top 10 Movie List." *The Conversation*. Last modified December 10, 2014. Accessed March 12, 2016. https://theconversation.com/bad-hobbits-die-hard-how-to-make-a-better-top-10-movie-list-35247.

Dibbets, Karel. *Film in The Netherlands from 1896*. Accessed March 12, 2016. http://www.cinemacontext.nl/.

Dinshaw, Carolyn. *How Soon Is Now? Medieval Texts, Amateur Readers, and the Queerness of Time*. Durham: Duke University Press, 2012.

Frame, Janet. *In the Memorial Room*. Melbourne: text publishing, 2013.

Gibbs, Frederick W. and Trevor J. Owens. "Hermenuetics of Data and Historical Writing." In *Writing History in the Digital Age*, edited by Jack Dougherty and Kristen Nawrotzki. Ann Arbor: University of Michigan Press/Digital Culture Books (Online), 2013. http://writing-history.trincoll.edu/data/gibbs-owens-2012-spring/.

Hitchcock, Tim. "Academic History Writing and the Headache of 'Big Data.'" *Historyonics*. Last modified January 30, 2012. Accessed March 12, 2016. http://historyonics.blogspot.com/2012/01/academic-history-writing-and-headache.html.

"Imagination in the Light of Nature (Terence McKenna)." YouTube video, 1:37:13, from a presentation at the Earth Trust Benefit, October 2nd, 1992. Posted by "Deus Ex Mckenna ~ Terence Mckenna Archive." July 9, 2011. https://youtu.be/IERPnScenBM?t=8m14s.

Koch, Peter-Paul. *Fluid Thinking*. Accessed March 12, 2016. http://www.digital-web.com/articles/fluid_thinking/.

Maltby, Richard, Dylan Walker, and Mike Walsh. "Digital Methods in New Cinema History." In *Advancing Digital Humanities: Research, Methods, Theories*, edited by Paul Arthur and Katherine Bode, 95–112. Hampshire: Palgrave Macmillan, 2014.

Maltby, Richard, et al. *Australian Cinemas Map*. Accessed March 12, 2016.

http://auscinemas.flinders.edu.au/

Maltby, Richard. "New Cinema Histories." In *Explorations in New Cinema History: Approaches and Case Studies*, edited by Richard Maltby, Daniel Biltereyst, Philippe Meers, 3–40. Oxford: Wiley-Blackwell, 2011.

Pauwels, Luc. *Visual and Digital Cultures Research Centre (ViDi)*. Accessed March 12, 2016. https://www.uantwerpen.be/en/rg/vidi/projects-and-publica/projects/.

Tansel, Abdullah Uz. "Temporality in Databases." In *Advances in Information Systems*, edited by Tatyana Yakhno, 1–3. Berlin: Springer-Verlag Berlin Heidelberg, 2004.

"The *Economist* Explains: The Backlash against Big Data." *Economist*. Last modified April 20, 2014. Accessed March 12, 2016. http://www.economist.com/blogs/economist-explains/2014/04/economist-explains-10.

Treveri Gennari, Daniela, et al. *Italian Cinema Audiences*. Accessed March 12, 2016, http://italiancinemaaudiences.org/.

Verhoeven, Deb. "New Cinema History and the Computational Turn." In *WCCA 2012: "Beyond Art, Beyond Humanities, Beyond Technology: A New Creativity,"* edited by Claudio da Rocha Brito, Melany M. Ciampi, Luis Amaral, and Rosa Vasconcelos. Guimarães: COPEC-Science and Education Research Council, 2012.

Verhoeven, Deb, Alwyn Davidson, and Bronwyn Coate, "Australian Films at Large: Expanding the Evidence about Australian Cinema Performance." *Studies in Australasian Cinema* 9, no.1 (2015): 7–20.

Verhoeven, Deb, Colin Arrowsmith, and Alwyn Davidson. *Cinema and Audiences Research Project*. Accessed March 12, 2016. http://caarp.edu.au/.

Verhoeven, Deb, Colin Arrowsmith, Bronwyn Coate, and Alwyn Davidson. *Cinema Cities Index*. Accessed March 12, 2016. http://cinemacities.com/.

———. *Film Impact Ratings*. Accessed March 12, 2016. http://www.reelmeasures.com/.

HOW IS A DIGITAL PROJECT LIKE A FILM?

Miriam Posner

Does classical film theory have anything to teach us about digital humanities? Formulated this way, the question actually retraces my own intellectual trajectory. Trained in a fairly conservative film studies PhD program, I now work professionally as a digital humanities scholar. My dissertation, on medical films, didn't exactly dwell on the canon, but it wasn't a digital project either. When people ask me how I got from one place to the other, it can be challenging to try to reconstruct the through-line.

There is a connection, though; in fact, film theory inflects a lot of my digital humanities work, both the work I'm able to do now and my ambitions for the kind of work I might someday do. In particular, digital humanities work prompts me to reflect frequently on one of the most absorbing, enduring questions in film studies: the problem of how individual slices of time and space become a narrative. How does a series of frames take on the power to tell a story, and (if successful) to mobilize people's emotions? How can a film tell stories that do justice to the reality and complexity of the world as people experience it? Must a filmmaker, if he or she is to argue for a different way of seeing the world, abandon conventional narrative altogether?

Perhaps you can anticipate the connections here to databased digital projects. As the frame is the basic unit of the photographic film, so we might see the database row as the fundamental unit of the digital project. To immediately hedge, this statement is, on its face, both right and wrong. An individual frame, as a film scholar certainly knows, is highly constrained and carefully constructed; it is an artificial view of the way things actually appeared. In this way, it's not unlike data as the digital humanist constructs and experiences it: subjective, carefully framed, excluding as much as

it reveals.

But the photographic image contains a plenitude and indeterminacy of information and meaning that structured data can't offer. The photograph, whatever it purports to be, is in fact "a message without a code," Barthes tells us, "a continuous message."[1] Is the subject smiling or grimacing? What is that in the background? What lies just outside the frame, and is that the photographer's reflection in the window? Structured data, on the other hand, is discrete by definition. The boundaries of an individual datum are clear, and every practice of database design—controlled vocabularies, authorities, interchangeable schema—works to eliminate ambiguity. Data also lacks direct indexicality, a fundamental property of the photographic image as classical film theory understands it. For André Bazin and later neorealist film theorists, indexicality was a near-mystical quality of mechanical (or photochemical) images of reality. The light embedded in the photographic image connects the photograph physically with the moment it has captured, thereby establishing an inextricable link between the two. The photograph, by virtue of the physical thread (however slender) that connects it to a specific time and place, maintains a tie to reality that offers the potential for accident, surprise, and reading against the grain.[2] "The photographic image is the object itself," André Bazin famously wrote, "the object freed from the conditions of time and space that govern it."[3]

Data, because it is interpreted and lifted from the phenomena it describes, maintains no such physical connection, in the Bazinian sense. We know that data describes an object only because the data-creator tells us this is so. A dataset stands at several levels removed from its subject. (All the more strange that "data" is our current moment's emblem of serious-minded objectivity!)

So while the dataset and the photograph both claim to represent reality, the photograph (as described in classical film theory)

maintains a direct, physical relationship to its subject that a dataset, by definition, cannot. The two modes of representing information, then, have fundamental ontological differences. And yet, there are important similarities in the way these two methods of dividing time and space are then strung together to create knowledge. It is here that film theory has the most to say: in the essential problem of assembling discrete, fungible units of information into a whole that claims to represent truth.

Before we continue too far in this vein, we should consider to what extent filmic theories of narrative can reasonably be applied to digital projects. Here, I want to expand on the best-known work on database and narrative, which takes up the question of whether the two modes are naturally antagonistic or complementary. Lev Manovich, in "Database as Symbolic Form," argues that the database is poised to overtake the narrative as the form best suited to expressing the contemporary condition. Indeed, he argues, this shift has already taken place: "Many new media objects do not tell stories; they don't have beginning or end; in fact, they don't have any development, thematically, formally or otherwise which would organize their elements into a sequence."[4] Katherine Hayles, in contrast, calls narrative and database "symbionts": two separate species which nevertheless depend on each other for their existence.[5] We accumulate isolated, parameterized slices of information in databases, Hayles argues, but these facts only acquire meaning when retrieved from databases and strung together in causal chains, which is to say, interpreted by human beings as narratives.

To adapt Hayles's argument to the digital humanities context, I rely on a definition of a digital project that I've previously outlined, in an argument that draws on Johanna Drucker and David Kim's DH101 coursebook.[6] The database is part of a digital project, I argue, but only part: a full analysis of a digital scholarly project requires a consideration of the sources being examined, the ways

in which those sources are transformed into data, and the ways in which that data is then presented to the viewer.

In a typical digital project built on a database, structured data is stored in a database in the form of linked tables. Using a scripting language like PHP or Javascript, an interface designer can retrieve data from the database and then format it with the markup language HTML and the styling language CSS. You have used a databased platform like this one, perhaps unwittingly, if you have posted to a WordPress blog, Tumblr, or even Facebook.[7] Most web users don't encounter data as it exists in a database but as it is retrieved, formatted, and presented on a website.

It is at this top level, the level of presentation, or interface, that the viewer generally comes into contact with the work (if the project is consumed as intended). Because, as Hayles argues, data has little human-comprehensible meaning when confined in a database, the author of a digital project must make that data legible in some way through an interface.[8] The presentation layer of a digital project creates its own constraints and possibilities: it is a "zone of affordances," in Johanna Drucker's formulation.[9] An interface without content makes no sense, just as a database without a retrieval mechanism is quite difficult to understand. It's the interaction of the two formats that produces the knowledge-effect that the viewer experiences.

So can we usefully think of that hybrid knowledge-object as a kind of narrative? Whatever their provenance, narratives must, in order to earn this classification, display a certain minimal set of qualities. At their most basic level, narratives are causal chains; one must be able to impute cause and effect to the information presented.[10] Interfaces to digital projects do not necessarily bind one piece of data to the next; indeed, many projects take advantage of the possibility for juxtaposition and remixing afforded by a database. Scalar, for example, the content-management system

for the presentation of long-form digital works, affords the arrangement of "pages" into narratives that the reader can then reorder and explore at will.[11]

And yet, that process of exploration and reordering is itself constrained by Scalar's zone of affordances. There is, of course, the crucial question of which data is included in the database at all. But that aside, Scalar's possibilities for reordering and remixing are not in fact endless; one can't, for example, layer pages on top of each other, loop among pages, cut media in half, or, really, individuate the appearance of individual pages to any great extent.

This, I think, is not a defect in Scalar's design, but an acknowledgment of some of the basic tenets of human meaning making. We require causal chains in order to develop meaning out of any collection of facts, and—despite the Web's apparent potential for radical juxtaposition—we seem to like stories that are built (like film) out of "segments of the same order of thing."[12]

Comparisons to film come in handy here in helping us to understand the implications of this flattening effect. Digital projects don't necessarily join one data point to the next, but in flattening reality into machine-readable data, they force the world into interchangeable parts (much as frames can be interpolated into a filmstrip). In narrowing the possibilities of expression to a zone of affordances dictated by the interface, digital projects also constrain the universe of meaning that can be derived from any dataset, much the way the selection and juxtaposition of frames place limits on the story an individual film tells. We want to believe that a platform like Scalar offers unprecedented possibilities for making meaning of data, but we find, in practice, that legibility dictates that we harness information together in a logical chain of events, that most quotidian of forms.

Film theory is good at capturing this peculiar brand of agony: the

doomed effort to represent the diversity of human experience in a medium that imposes inflexible constraints. What to do about this impossible task? film theorists have asked. Should one assiduously work to capture events as they unspool, regardless of how futile the effort to record everything will ultimately be? Or should one intervene in narrative conventions in order to provoke and disturb the viewer? Here, we might contrast the approach of American cinéma vérité filmmakers like D. A. Pennebaker with that of an experimental documentarian like Trinh T. Minh-Ha. In the former, the filmmaker makes a faithful effort to depict reality as it unfolds; in the latter, the filmmaker explodes any possibility of narrative. To draw a comparison with the databased digital project, should one collect ever more and various data to create the most truthful representation of human experience? Or should one intervene in the form of the database, interface, or algorithm itself, sacrificing legibility to radical rule bending?

To make this question more concrete, compare Jennifer Terry and Raegan Kelly's "Killer Entertainments" with *Digital Harlem: Everyday Life 1915–1930*.[13] Terry and Kelly's piece, which presents viewers with combat footage from the Iraq War, is baffling, even upsettingly so. Three videos play simultaneously, their soundtracks overlapping as keywords ("first person POV," "B roll") drift onto the frame. It's hard to understand what one's seeing, and virtually impossible to put these media in any kind of order. In contrast, *Digital Harlem* presents us with an interface that feels comfortingly familiar, in part because it's built atop a technology, Google Maps, that many of us use every day. In *Digital Harlem*'s schema, time is divided neatly into decades and then subdivided into years, and data can be mapped and filtered according to the kind of thing it is: an event, person, or place, and then, more deeply, according to whether it's a man or woman, barbershop or bowling alley, drug deal or fashion show.

Digital Harlem puts an impressive amount of data at the viewer's

disposal, but even as it heaps data point upon data point, it seems to suffer beneath the pressing question of *what kind of thing this is*. For in what world, really, does an "Abortion" belong to the same order of things as a "Murder (of Spouse)"? Whose gender is so neatly demarcated as male, female, or unknown? Who experiences time as a clicking forward of decades, rather than the unspooling of life? Is this, finally, a more truthful, useful, or generative casting of history than the "radically decontextualized" sensorium Terry offers in "Killer Entertainments"?[14]

Digital humanists (Terry aside) seem startled at the discovery that the database imposes impossible constraints on the depiction of human experience. But film scholars have long known that it is absurd to attempt to capture human experience in a photographic narrative. Because we understand the photographic image—its trickery, its inherent limitations, the world beyond its frame—we understand how essentially false is any work's claim to represent "reality" in all its plenitude and contingency. To argue that a film is fully representative of any given event is to be unforgivably naïve; we know that every work is constructed, no matter how transparent it appears.

But somehow we feel there's something valiant in the attempt to capture human experience, even in these inadequate media. Writing on Rossellini's *Paisà*, André Bazin observed that the film's essential unit is not the shot but the "fact," one slice of time and space, itself worthy of interpretation and filled with meaning. *Paisà* is rife with gaps and omissions, but so much the better: "The mind has to leap from one event to the other as one leaps from stone to stone in crossing the river."[15] The best films are beautiful not because they claim earnestly to represent reality, but because they acknowledge this feat's impossibility but *keep trying anyway*, honoring their viewers by trusting them to make their way from stone to stone.

There's a potential for a digital humanities that holds toward data the same vexed, impossible loyalty with which media scholars honor the photographic image. In this version of digital humanities, scholars would view data neither as fully adequate to reality nor as necessarily mendacious, but as one moment, a slice of time and space. The best work would not be the most comprehensive—just as the best films are not the most verisimilitudinous—but that which exhibits the most sophistication, the most humanity, in making the leap from fact to narrative.

I don't think digital humanities is there yet, but I think this is an opportunity for media scholars. This is why I think the best possibilities for the intersection of digital humanities and media studies lie not so much in counting frames or automating facial recognition (though this is interesting in its way) as by bringing to digital humanities the peculiar agony of the media scholar: the belief, simultaneously, that all stories are lies and that there's truth in their telling.

ENDNOTES

1 Roland Barthes, "The Photographic Message," in *A Barthes Reader*, ed. Susan Sontag (New York: Hill and Wang, 1982), 196.

2 This reading of "indexicality" is distinct from Charles Sanders Peirce's semiotic sign system, in which the index points at its referent but does not necessarily contain an existential link to it. You might think of this difference as akin to the distinction between a pointing finger and smoke from a fire: one points, while the other contains elements of the phenomenon that produced it. It's not unusual to call the Bazinian use of the term (the "smoke-from-a-fire" definition) a misreading of Peirce, but whether it's a misreading or a different concept altogether, this use of indexicality has influenced a great deal of classical film theory, and I find it quite useful in understanding why photographs have such persuasive ability. See Charles Sanders Peirce, "What Is a Sign?," in *The Essential Peirce, Volume 2: Selected Philosophical Writings (1893–1913)*, ed. Peirce Edition Project (Bloomington: Indiana University Press, 1998); Martin

Lefebvre, "The Art of Pointing. On Peirce, Indexicality, and Photographic Images," in *Photography Theory*, ed. James Elkins (New York; London: Routledge, 2007), 220–44.

3 André Bazin and Hugh Gray, "The Ontology of the Photographic Image," *Film Quarterly* 13, no. 4 (Summer 1960): 7.

4 Lev Manovich, "Database as Symbolic Form," *Convergence: The International Journal of Research into New Media Technologies* 5, no. 2 (June 1, 1999): 80, doi:10.1177/135485659900500206.

5 N. Katherine Hayles, "Narrative and Database: Natural Symbionts," *PMLA* 122, no. 5 (October 1, 2007): 1603–08.

6 Miriam Posner, "How Did They Make That? The Video!," *Miriam Posner's Blog*, last modified April 17, 2014, accessed March 6, 2016, http://miriamposner.com/blog/how-did-they-make-that-the-video/; Johanna Drucker and David Kim, "Introduction to Digital Humanities | Concepts, Methods, and Tutorials for Students and Instructors," UCLA Center for Digital Humanities, last modified 2013, accessed March 6, 2016, http://dh101.humanities.ucla.edu/.

7 More recently, database designers have made use of document-oriented (as opposed to relational) databases such as NoSQL and MongoDB. In the document-oriented database, data is described within documents themselves, not stored in rows as in a traditional, SQL database. However, my key point here is not about the literal grid structure of the relational database table, but about the parameterization and circumscription intrinsic to the transformation of any information into computable data.

8 It's not exactly true that an individual piece of data has little or no meaning on its own. Even the simplest statement, such as "Director: Agnès Varda" is, in fact, human-comprehensible, and might be subjected to all manner of scrutiny and meaning making. I'm concerned here, though, with what happens when units of information like this one get heaped on top of each other in great piles, as in a database. How are we to make sense of all of this, and is there any part of this process of information retrieval that can be usefully compared to narrative?

9 Johanna Drucker, "Humanities Approaches to Graphical Display," *Digital Humanities Quarterly* 5, no. 1 (2011), http://digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

10 Hayden White, "The Value of Narrativity in the Representation of Reality," *Critical Inquiry* 7, no. 1 (October 1, 1980): 5–27.

11 http://scalar.usc.edu/.

12 J. Drucker, "Humanities Approaches to Interface Theory," *Culture Machine* 12 (2011): 12–13.

13 Jennifer Terry, "Killer Entertainments." *Vectors Journal of Culture and Technology in a Dynamic Vernacular* 3, no. 1 (Fall 2007); *Digital Harlem: Everyday Life 1915–1930*, University of Sydney. Last modified 2010. Accessed March 6, 2016. http://digitalharlem.org/.

14 "Author's Statement," in Terry, "Killer Entertainments."

15 André Bazin, "An Aesthetic of Reality: Cinematic Realism and the Italian School of Liberation," in *What Is Cinema? Volume 2*, trans. Hugh Gray (Berkeley and Los Angeles: University of California Press, 2005), 35.

BIBLIOGRAPHY

Barthes, Roland. "The Photographic Message." In *A Barthes Reader*, edited by Susan Sontag, 194–210. New York: Hill and Wang, 1982.

Bazin, André and Hugh Gray. "The Ontology of the Photographic Image." *Film Quarterly* 13, no. 4 (Summer 1960): 4–9.

Bazin, André. "An Aesthetic of Reality: Cinematic Realism and the Italian School of Liberation." In *What Is Cinema? Volume 2*, translated by Hugh Gray, 16–40. Berkeley and Los Angeles: University of California Press, 2005.

Digital Harlem: Everyday Life 1915–1930, University of Sydney. Last modified 2010. Accessed March 6, 2016. http://digitalharlem.org/.

Drucker, Johanna. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly* 5, no. 1 (2011). http://digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

Drucker, Johanna. "Humanities Approaches to Interface Theory." *Culture Machine* 12 (2011): 1–20.

Drucker, Johanna and David Kim. "Introduction to Digital Humanities | Concepts, Methods, and Tutorials for Students and Instructors," UCLA Center for Digital Humanities. Last modified 2013. Accessed March 6, 2016. http://dh101.humanities.ucla.edu/.

Hayles, N. Katherine. "Narrative and Database: Natural Symbionts."

PMLA 122, no. 5 (2007): 1603–08.

Lefebvre, Martin. "The Art of Pointing. On Peirce, Indexicality, and Photographic Images." In *Photography Theory*, edited by James Elkins, 220–44. New York: Routledge, 2007.

Manovich, Lev. "Database as Symbolic Form." *Convergence: The International Journal of Research into New Media Technologies* 5, no. 2 (1999): 80–99. doi:10.1177/135485659900500206.

Peirce, Charles Sanders. "What Is a Sign?" In *The Essential Peirce, Volume 2: Selected Philosophical Writings (1893–1913)*, edited by the Peirce Edition Project, 4–10. Bloomington: Indiana University Press, 1998.

Posner, Miriam. "How Did They Make That? The Video!" *Miriam Posner's Blog*. Last modified April 17, 2014. Accessed March 6, 2016. http://miriamposner.com/blog/how-did-they-make-that-the-video/.

Terry, Jennifer. "Killer Entertainments." *Vectors Journal of Culture and Technology in a Dynamic Vernacular* 3, no. 1 (Fall 2007).

White, Hayden. "The Value of Narrativity in the Representation of Reality." *Critical Inquiry* 7, no. 1 (1980): 5–27.

PART III: **ANALYZING IMAGES, SOUNDS, WORDS**

CODING AND VISUALIZING THE BEAUTY IN HATING MICHELLE PHAN: EXPLORATORY EXPERIMENTS WITH YOUTUBE, IMAGES, AND DISCUSSION BOARDS

Tony Tran

With the arrival of YouTube's tenth anniversary in April 2015, much attention was given to how a decade of Charlie biting a finger, a chemical reaction between Mentos and Coke, a song about Friday, and Chinese teenagers lip-syncing in a dorm all influenced the ways in which we watch, share, and understand popular culture in an increasingly digital world. While the viral nature of YouTube and its own discourses have often emphasized the value of both the now and the future of the internet, YouTube's anniversary also allows us as media scholars to reconsider and reflect on how to study a decade of online audiovisual broadcasting and storage. As several scholars have argued, YouTube's immense popularity has created various forms of public and informal collections and archives that have allowed us greater access to overwhelming amounts of moving image media and data, including materials that extend beyond YouTube.[1] This abundance of media is not a new problem; scholars of news media, fandoms, and soap operas and other long-running television series have always grappled with the issue of designating a manageable scope of study. But YouTube and other online sites offer scholars greater opportunity to use software to download, manipulate, organize, and pre-analyze large digital media and datasets. In turn, this capability raises questions of how we can approach and harness digital media when it is abundant, nonlinear, and diverse, and how we can begin to incorporate it to aid current methodologies in media studies.

Through experimenting with text-mining software, ImageJ, and jQuery scripts, I explore and reflect upon the possibilities, limits, and frustrations of working with multimedia data produced by

makeup "guru" Michelle Phan's YouTube channel and a message board forum responding to her videos. With these two datasets, I outline the technical processes of coding, ripping, formatting, and visualizing media related to Michelle Phan. Reflecting on these processes, I argue that while the results of these experiments are limited, these datasets and processes do provide an intriguing environment where we can further interrogate the methods and assumptions we employ to discover, organize, and analyze media texts and audiences.

This essay mainly focuses on methodology rather than showcasing and elaborating computational results and arguments about Phan and her audience. In other words, while I embarked on these experiments to learn more about Michelle Phan, I am ultimately more interested in how these experiments with coding and visualization can influence media scholarship. Additionally, this essay illustrates a low-budget (read: zero), one-person project to demonstrate that mostly self-taught coding can produce valuable results. With this in mind, these experiments admittedly contain dead ends and unanswered questions, as well as showcasing my own technical limits in producing efficient coding and workflows. As Mark Williams stated at the 2015 Arclight Symposium, digital humanities entails everyone stepping outside of their comfort zone in some manner, and this essay is hopefully a healthy and transparent example of the discomfort we sometimes experience. In positioning these processes as exploratory experiments with varying levels of success, I follow Matthew Kirschenbaum's framing of digital humanities, where computational methods within the digital humanities are best understood in terms of starting conversations or "provocation" not problem solving. For Kirschenbaum, the results of computational methods are not *the* end point, but instead represent the possibility of generating parallel evidence and additional starting points that acknowledge the richness of human culture and where new exploratory processes can be created to better understand digital cultures.[2]

STUDYING MICHELLE PHAN AND HER "ANTI-PHANS"

An early adopter of YouTube and vlogging (video blogging), Vietnamese American Michelle Phan is a self-taught makeup artist who is now one of the most-subscribed-to women on YouTube. Her channel consists of a collection of over 380 makeup and fashion tutorials that map her transition from an unknown Ringling College of Art and Design student in 2007 to an influential Lancôme-sponsored cosmetic guru with over 8.1 million subscribers and 1.2 billion views.[3] Seeking to promote the concept of female empowerment through makeup, Phan's tutorials range from five to fifteen minutes and employ voice-overs to narrate instructions as she applies cosmetic products to herself. Tutorial themes include broad topics like "Simple, Everyday Looks" or "Makeup for Glasses," as well as specific videos on how to look like a K-Pop Star, Zombie Barbie, or Lady Gaga. With her rise in subscribers, she has recently released her own cosmetic line, EM, as well as venturing into other business avenues such as starting a media production company.[4]

While Phan's numbers do indicate her wide popularity, within these subscribers and viewers are a subset of passionate viewers I label Anti-Phans. An embellishment of Jonathan Gray's concept of the anti-fan, which highlights how hate or dislike of a text/object can produce intense passion and communities similar to fandom, Anti-Phans simply hate Michelle Phan.[5] But like most fandoms and anti-fandoms, the relationships between audiences and texts are complex, and for Anti-Phans, their hate equally expresses a detailed knowledge of Phan, her videos, and her growing brand. A few hours after Phan has uploaded a video, a new thread is created on GuruGossiper.com's "Trash a Guru" forum to discuss the video, with Anti-Phans watching, deconstructing, and critiquing every detail. Even though hatred is a key emotion, Anti-Phans are passionate about consuming Phan's videos and follow her career with an intense fervor, producing over 400 topic posts and 50,000 multimedia comments.[6]

Before applying software processes to these two main datasets—Phan's YouTube video collection and the Anti-Phan message board—I wanted to first get a sense of them through traditional qualitative methods of analysis, including using textual and discourse analysis by watching/skimming about seventy-five videos and their corresponding threads. Broadly speaking, I am interested in how Phan constructs definitions of beauty, gender, class, and racialized bodies, specifically in Asian and Vietnamese American contexts, and how Anti-Phans negotiate these constructions. I concluded that Anti-Phans form their own identities and ideologies of beauty through the recognition of and resistance to Phan's consumerist notions of empowerment and beauty. Seeing Phan's deal with Lancôme as a corporate sellout, Anti-Phans (who often identify as former fans) remark how the "new" Phan conforms more with white/Western hegemonic notions and standards of gender and femininity to fit the perceived ideologies of the mainstream cosmetics industry. Additionally, they argue Phan capitalizes on and appropriates her Asian and Vietnamese identity as a way to sell products, rather than holding a deep commitment to her Vietnamese American cultural background. Overall, Anti-Phans contend that Phan's videos sell "female empowerment" through consumption while reinforcing harmful images of beauty and race in areas such as colorism, weight, and facial features.

While it is easy to position these Anti-Phans as active and progressive audiences, a closer look at their message boards indicates that the construction of these critical discourses about Phan also contain problematic definitions and binaries of beauty, the body, and race. For example, in attempts to counter Phan's hegemonic notions of beauty, Anti-Phans make harsh comments about how ugly Phan is in order to make their arguments through difference. In the process, Anti-Phans establish their own hierarchies of what is considered beautiful and reproduce the harmful ideologies they wish to critique. Furthermore, their discussion of Phan's race/ethnicity moves into essentialist terms, where Phan's mentioning of

her race becomes an illustration of "inauthentic" Vietnamese and Asian cultures and bodies.

In setting my textual parameters for a more focused close reading, my analysis included five videos and their corresponding message threads, including the Anti-Phan's comments. With this focus, I was able to produce a detailed and specific analysis, including greater attention to aspects such as word choice, vocal and written tones, audio-visual aesthetics (color, camera angles, music, etc.), editing, language structures, and the relationship between actors within specific cultural contexts. However, there is a possible—and admittedly reasonable—claim that my analysis is *too* specific due to the small size of the dataset/sample size. Are my analysis, arguments, and conclusions only applicable to these five videos and message threads, and if not, how can I show that my results can be applied to the larger Anti-Phan community? What possible larger trends am I missing in my analysis of such a small sample size? And from a practical standpoint, how can I efficiently produce and communicate close readings of relatively large amounts of data? In other words, what about the other 375 videos, 395 message board threads, and 49,500 comments? To move toward answering these questions, I wanted to use software to describe and visualize the content of several Michelle Phan videos and Anti-Phan responses in order to discover a larger context for my datasets. While I am unable to fully analyze all of the videos and threads, I was able to use software on 193 videos and 81 message board threads.

To start my computational analysis, I needed to acquire Phan's videos and convert them into formats I could manipulate. This process involved ripping 193 videos from YouTube using online sites like KeepVid.com and ClipConverter.cc. Inserting the YouTube URL link into these sites allowed me to download the videos for further processing. Before I could use these videos with ImageJ, the videos needed to be converted to images and

formatted and compressed to the same image and file size. Using Free Studio 5, a free video-to-JPEG image converter, the program went through each video and took a JPEG frame grab every two seconds, which, while arbitrary, gave a general view of each video without skipping over major details.[7] After converting all of the videos to images, I ended up with roughly 33,000 JPEG images, with each video having its own folder of images and organized in chronological order. Due to different cameras and aspect ratios, I used Free Picture Resizer to perform batch manipulations to make all of the images identical in size.[8] Once formatted, the images were imported into ImageJ as "stacks" to produce montages and z-projections.[9] For montages, ImageJ takes the imported images and chronologically places them across a grid.

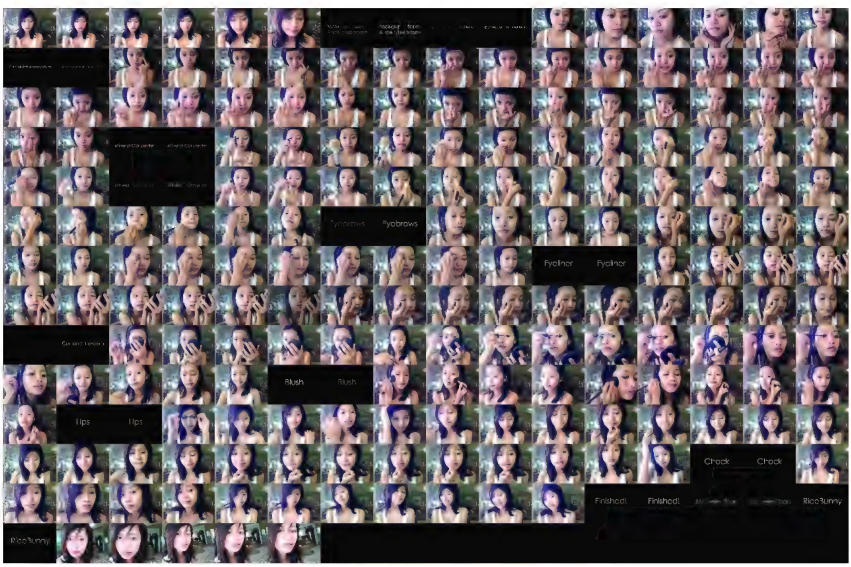


Figure 1. A montage of Phan's first video, "Natural Looking Makeup Tutorial."

The resulting montages gave me the chance to perform "distant readings" of Phan's videos.[10] Thinking of the montages as more of an exploratory tool, I was unconcerned with the specifics of content at this point; instead, I was focused on summarizing visual

content across 193 videos. By reducing videos into data that could be read with a single glance, "distant reading" allows for different and sometimes more efficient forms of comparing moving images, with the potential to see patterns across large datasets that can be obscured in close readings.[11] Instead of playing 193 tutorials at the same time, ImageJ gave me manageable options to analyze Phan's videos on a larger scale.

This does not mean "distant reading" is always productive. While ImageJ is able to make a montage of all 33,000 images, I did not find the results helpful. Although in other cases a montage of thousands of images may be revealing, seeing a single montage of over 33,000 screenshots was a bit overwhelming for my 14-inch laptop and seemed to obscure any discoveries.[12]

Figure 2. A montage of montages, or over 33,000 images of Michelle Phan Videos.

What was more useful were montages of individual videos. While I hesitate to argue that something major was discovered, I did find the montages were helpful in organizing and classifying videos.

This allows the researcher a quicker processing of the visual aspects of Michelle Phan's videos, which, as mentioned above, already number over 380. While I am not arguing this replaces watching the videos, it provides a broad view of the visual aesthetics of Phan's videos and information that allows for more selective viewing practices. For example, if I wanted to focus on videos specifically addressing eyes and/or eye makeup—a topic of interest for many Asian American Anti-Phans—a quick scan of the montages can help me make informed decisions about how to spend my research time. In many cases, a simple search of the titles may not register specific topics; ambiguous titles such as "Double Lines" might pass through a researcher's skim, but its montage shows a clear focus on eyes.

Figure 3. A montage of Phan's video, "Double Lines."

While it is possible to watch every video, this approach can be extended to projects with massive amounts of audio-visual data, and, by letting computers do some preliminary work through pre-analyzing processes, researchers can be more efficient in locating specific texts for deeper analysis, especially as hundreds of hours of new content are uploaded to YouTube every minute.

What really surprised me and highlighted new ways of looking at Phan's videos were the z-projections. To be honest, I initially produced z-projections just because I could with ImageJ. For z-projections, ImageJ takes the same images and instead of placing them on a grid to produce montages, it makes all the images the same opacity and stacks them on top of each other. While some z-projections produced blurry, abstract pictures, the majority of them are ghostly outlines of Phan's head (see figs. 4 and 5).

This revealed several insights I had not originally considered. First was the limited movement in the majority of her videos. If there was visual diversity, it would be more difficult to distinguish specific objects in the z-projections, but several of the images clearly show Phan's face (with several others being close-ups of her eyes). While I am unsure of how to apply this to a larger argument, the z-projections could be used to contextualize the five videos I selected in relation to the whole. It also raised the question: in what ways did Phan change after the Lancôme deal—a common Anti-Phans' accusation—if most of her videos are visually similar?

Another revelation was the visual presence of the Lancôme brand in Phan's videos (see fig. 4). As mentioned, many Anti-Phans argue that Phan is postfeminist in her promotion of empowerment through consumption, mainly through Lancôme products. When performing my qualitative analysis I, like several Anti-Phans, felt that the Lancôme brand became more visible within her videos. The z-projections suggest otherwise. Browsing through the images, the word Lancôme does appear, but it only visually registers for three videos out of 193—a very small percentage. While this evidence does not disprove Anti-Phan claims, I need to reevaluate how I frame this visibility in relation to her larger corpus.

Given the lack of Lancôme's visual branding, I began to wonder if the belief in Lancôme's growing role in Phan's tutorials was

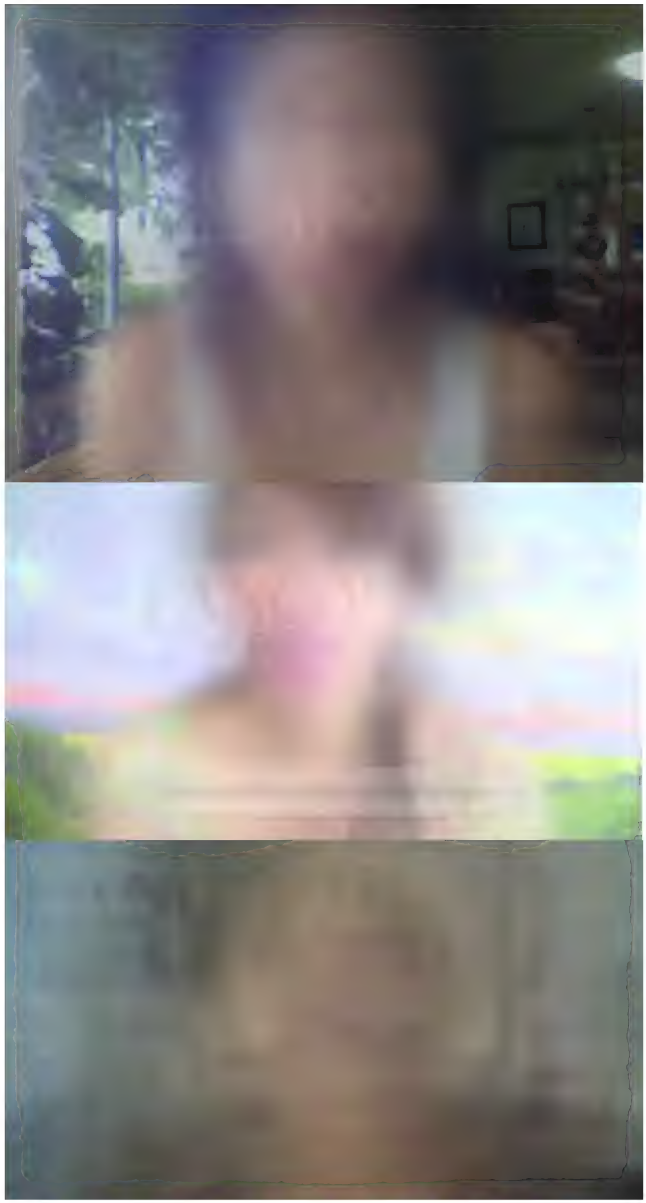


Figure 4. From top to bottom, z-projections of “Natural Looking Makeup Tutorial,” “How To’s For Perfume,” and “Day to Night Makeup.”

Figure 5. A montage of z-projections.

a result of her spoken words or the presentation of Lancôme's products—which would not register well on z-projections—rather than the direct use of its logo. I discovered that YouTube has a closed captioning system based on voice recognition that produces subtitles which can be downloaded as a SubRip Text (SRT) file. In addition to text, it also gives metadata about the time position of the text within the video:

```
00:00:14,880 --> 00:00:18,528
so don't beat yourself up over it it's perfectly
normal to have it
```

Here, Phan, discussing cellulite, speaks from the 14.880 to the 18.528 second mark. Whatever research potential this may have had, I soon discovered YouTube's closed captioning system did not work well. For instance, when Phan discusses model Cara Delevingne's style, Phan states "no conversation about Cara is complete without talking about the brows . . . unless you are blessed with

a full and healthy set of brows like Miss Delevingne." YouTube's closed captioning, however, produced this: "No conversation about car is complete without talking about routed unless are the last with a full in hockey centre proud like miss tel aviv."[13] The closed captioning system had about a 70% rate of error, leaving me searching for different methods to analyze Phan's words on a large scale.

More successful was the analysis of the message boards. To assemble this dataset, I downloaded the underlying HTML web code of the message boards and, using jQuery, I scrubbed the code to only collect the words and text within the comment boxes. Like most HTML, the code was structured and organized in a consistent manner:

```
<div class="content"><br /><br />One word
- EWW!<br /><br /><span style="font-weight:
bold"><span style="color: #FF0000">Video</span>
</div>
```

Breaking down the HTML structure, the information I wanted was found between specific tags. Using jQuery, I wrote a script that searched the HTML file for every element with a class of "content" and copied the text within this element into the variable "text."

```
<script>
var text = $("div.content").text();
document.write(text);
</script>
```

At the end of the script, jQuery printed all of the messages, which were then easy to move over to an Excel or text document. I organized the messages into two folders: one with the entire corpus of text in one file (7,219 comments/431,369 words) and the other

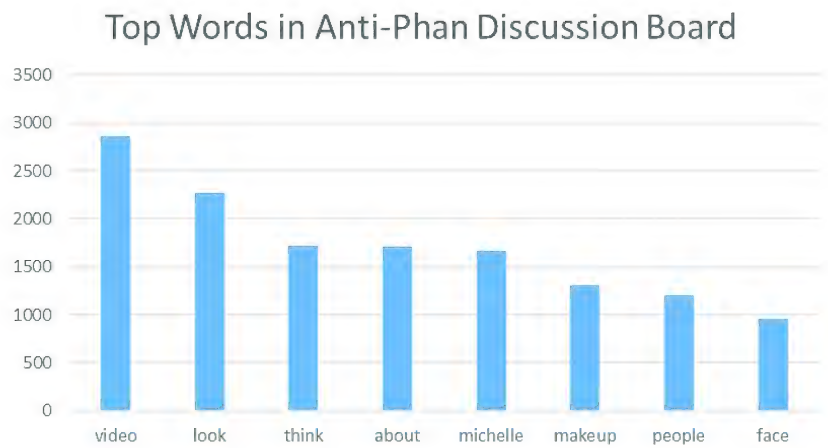


Figure 6. The top trending words within Anti-Phan posts.

separated into individual threads (81 in total). With these files, I processed them through the program TextSTAT (Simple Text Analysis Tool) to produce frequency counts of words appearing in these comments.[14] From the first file, which had all of the comments, I produced the graph below showing the most used words (after taking out commonly used words such as "the," "and," "I," "of," etc.).

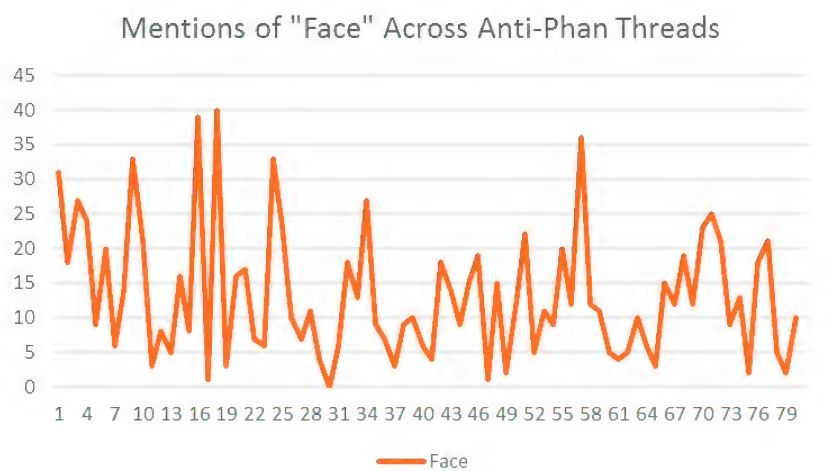


Figure 7. Mapping Michelle Phan's "Face" across Anti-Phan threads.

Initially, the graphing process produced underwhelming information. It was no surprise that a message board centered on Phan's makeup videos uses words like "video," "makeup," and "Michelle." Still, the results do provide some use for other words. For instance, the word "face" is one of the highest ranked words, which makes sense considering the amount of screen time Phan's face takes up in her videos. But its presence in this chart also illuminates and changes how we can interpret the deployment of that word in a user's comment. Take for instance TopCat's comment in one of the threads:

> Why does her face look so jacked up in this video O_O I never realized it till now, like her lips look waaay too plumped up, her chin is too long, her teeth are literally above each other.[15]

Instead of viewing this comment as a selective example or an outlier, the graph gives evidence and context that her face is a major topic of conversation within this specific group. Again, since this content analysis does not consider the tone or full context, it is difficult to clearly state that TopCat's discussion of Phan's "face" is representative of a common theme, or even if all instances of "face" refer to Phan's physical face. But by considering other qualitative (additional user comments) and quantitative (the montages and z-projections) evidence, we can begin to see how the word "face" trends and recurs as part of a larger movement and is more than just my selective choice.

Perhaps more useful is the addition of time to the equation. Using the separated threads and organizing them in chronological order, we can map out how and where "face" is used. With the Y axis representing the frequency, the X axis representing time, and each node/data point on the X axis representing one message board thread, we can begin to place specific quotes in relation to each other to see how words ebb and flow within a specific context. While this increases our understanding of how, where, and

when certain words are used, it also raises a lot of questions. If Phan's face is constantly present in her videos, why does the topic of her "face" vary over time so much? Why do certain videos and threads have little to zero mention of her face? Is this based on the aesthetics of the video or the users within the thread? Again, these experiments cannot fully answer these questions, but it can highlight these issues for qualitative methods to explore.

Increasing the variables involved can also result in producing new and interesting questions. Moving beyond one word, I broke the "face" into several subjects, such as "eyes," "lips," "nose," and "chin." Similar to "face," I mapped these words over time and layered them on top of each other to produce a comparative graph.

While the result is a bit overwhelming, it also indicates which threads are focused on certain aspects of Phan's body, raising the questions of why some topics appear more frequently in certain threads and how these topics relate to each other.

Here, it might seem I am limiting the object of study by introducing "fixed" categories such as "eyes" or "lips." However, the resulting graph is not presented as an answer, but rather as a tool or product open to further inquiry. In this case, I am not stating that the graph produced *is* "eyes" in its entirety, or even how it is used. What I wish to illustrate is *where* these topics are being discussed so that I can use qualitative methods to explore how these terms are being negotiated and (re)defined. In other words, by using both quantitative and qualitative methods, we can recognize language is often in the form of static categories (in this case digital text), but its social and cultural meanings are constantly in flux.

CONCLUSION

I would like to conclude with a brief discussion of how these exploratory experiments have helped me to become more reflexive in my own work. As mentioned above in the discussion of

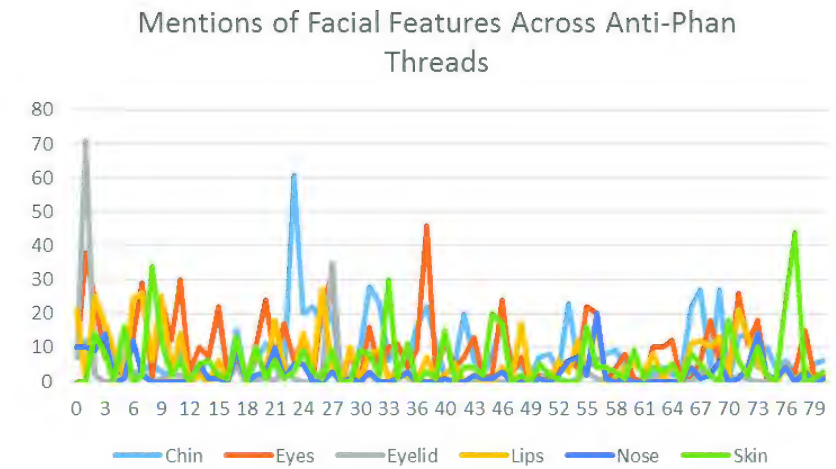


Figure 8. Phan's facial features.

z-projections, I was mistaken about how much (or rather in what manner) the Lancôme brand had become visible. For me, this discovery has raised questions of what biases and standpoints I have while doing qualitative analysis. My original argument about Phan and her Anti-Phans showcased some of my ambivalence toward each side, as I was critical of how they were constructing dueling definitions of beauty standards, both of which had potentially problematic implications. However, I will admit to being more partial to the Anti-Phans, who in many cases present a more thoughtful and inclusive discourse on beauty and gender. Was I influenced too much by the discourses of Anti-Phans to see resistance against issues that did not exist? I do not wish to disregard these comments that critiqued Phan's relationship with Lancôme on the basis of some "Truth" I discovered using computational methods. Still, while I feel the resistance presented here against large corporations and capitalism is a progressive sign, I think I should consider how I came to the conclusion that Lancôme was becoming more visible in Phan's videos and question if I am targeting types of resistance that may be overemphasized, thus misrepresenting a community. In a similar vein, the larger con-

texts of my evidence require fuller consideration. What does it mean if a user employs words that are not commonly used within the larger community, or if I focus on a video that is stylistically different from the other 380? The blurring of the boundaries between traditional and digital methods raises several questions, but, in the case of researchers, this productively allows us to build and evolve what and how we ask questions.

ENDNOTES

1 See Karen Gracy, "Moving Image Preservation and Cultural Capital," *Library Trends* 56, no. 1 (2007): 183–97; Robert Gehl, "YouTube as Archive: Who Will Curate this Digital Wunderkammer?" *International Journal of Cultural Studies* 12, no. 1 (2009): 43–60; Pelle Snickars and Patrick Vonderau, eds., *The YouTube Reader* (Stockholm: National Library of Sweden, 2009); Ryan Skinnell, "Circuitry in Motion: Rhetoric(al) Moves in YouTube's Archive," *Enculturation* 8 (2010): accessed November 15, 2015, http://enculturation.net/circuitry-in-motion.

2 Matthew Kirschenbaum, "The Remaking of Reading: Data Mining and the Digital Humanities," *National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation* 9 (2009): accessed November 15, 2015, http://www.csee.umbc.edu/~hillol/NGDM07/abstracts/talks/MKirschenbaum.pdf.

3 "Michelle Phan YouTube Homepage," accessed November 20, 2015, https://www.youtube.com/user/MichellePhan.

4 Born in Boston, Phan first started blogging as a teenager on early online platforms such as Xanga and Asian Avenue. She then started making video tutorials with a webcam in her parents' living room. A key turning point was when her Lady Gaga makeup tutorial went viral in 2010, which drew attention from both YouTube and cosmetics companies. Afterwards, she transitioned to more professional media equipment and began expanding her brand to include fashion, life tips, traveling segments, and her own makeup line. For a longer autobiography, see Michelle Phan, *Make Up: Your Life Guide to Beauty, Style, and Success—Online and Off* (New York: Harmony Books, 2014).

5 Jonathan Gray, "Antifandom and the Moral Text: Television

without Pity and Textual Dislike," *American Behavioral Scientist* 48, no. 7 (2005): 840–58.

6 "Michelle Phan TRASH A GURU message board," accessed November 20, 2015, http://gurugossiper.com/index.php. During this analysis, this board was available to the public; however, a (free) login is now required to see the board.

7 "Free Studio Download Page," accessed November 20, 2015, http://www.dvdvideosoft.com/free-dvd-video-software.htm#.Vd_XqvlViko.

8 "Free Picture Resizer Download," accessed November 20, 2015, http://download.cnet.com/Free-Picture-Resizer/3000-12511_4-10297789.html.

9 For full documentation for ImageJ, see "ImageJ User Guide," accessed November 20, 2015, http://rsb.info.nih.gov/ij/docs/guide/146.html.

10 See Franco Moretti, "Conjectures on World Literature," *New Left Review* 1 (2000): 54–68.

11 For more examples and discussion of visualizing large sets of images, see Lev Manovich, "Media Visualization: Visual Techniques for Exploring Large Media Collections," in *The International Encyclopedia of Media Studies*, Volume VI: *Media Studies Futures*, edited by Kelly Gates (London: Blackwell, 2012), accessed November 20, 2015, http://manovich.net/index.php/projects/media-visualization-visual-techniques-for-exploring-large-media-collections; "Software Studies Initiative Website," accessed November 20, 2015, http://lab.softwarestudies.com/.

12 For examples of useful large montages, see Software Studies Initiative projects "PowerWall Presenter: using gigapixel displays in humanities and social sciences," accessed November 20, 2015, http://lab.softwarestudies.com/2008/04/powerwall-new-presentation-strategies.html; "Selfiecity Project Website," accessed November 20, 2015, http://selfiecity.net/; "On Broadway Project Website," accessed November 20, 2015, http://www.on-broadway.nyc/.

13 "Cara Delevingne Makeup," YouTube video, 4:09, posted by "Michelle Phan," September 6, 2013, https://www.youtube.com/watch?v=D7elkw50rcg.

14 "TextSTAT – Simple Text Analysis Tool," accessed November 20, 2015, http://neon.niederlandistik.fu-berlin.de/en/textstat.

15 "Guru Gossip Board – [Video] a Guy, a Girl," accessed November 20, 2015, http://gurugossiper.com/viewtopic.php?f=72&t=9203&p=585529&hilit=face+look+so+jacked+up#p585529.

BIBLIOGRAPHY

"Cara Delevingne Makeup." YouTube video, 4:09. Posted by "Michelle Phan." September 6, 2013. https://www.youtube.com/watch?v=D7elkw50rcg.

Gehl, Robert. "YouTube as Archive: Who will Curate this Digital Wunderkammer?" *International Journal of Cultural Studies* 12, no. 1 (2009): 43–60.

Gracy, Karen. "Moving Image Preservation and Cultural Capital." *Library Trends* 56, no. 1 (2007): 183–97.

Gray, Jonathan. "Antifandom and the Moral Text: Television without Pity and Textual Dislike." *American Behavioral Scientist* 48, no. 7 (2005): 840–58.

Kirschenbaum, Matthew. "The Remaking of Reading: Data Mining and the Digital Humanities." *National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation* 9 (2009). Accessed November 15, 2015. http://www.csee.umbc.edu/~hillol/NGDM07/abstracts/talks/MKirschenbaum.pdf.

Manovich, Lev. "Media Visualization: Visual Techniques for Exploring Large Media Collections." In *The International Encyclopedia of Media Studies*, Volume VI: *Media Studies Futures*, edited by Kelly Gates. London: Blackwell, 2012. Accessed November 20, 2015. http://manovich.net/index.php/projects/media-visualization-visual-techniques-for-exploring-large-media-collections.

Moretti, Franco. "Conjectures on World Literature." *New Left Review* 1 (2000): 54–68.

Phan, Michelle. *Make Up: Your Life Guide to Beauty, Style, and Success—Online and Off*. New York: Harmony Books, 2014.

Skinnell, Ryan. "Circuitry in Motion: Rhetoric(al) Moves in YouTube's Archive." *Enculturation* 8 (2010). Accessed November 15, 2015. http://enculturation.net/circuitry-in-motion.

Snickars, Pelle and Patrick Vonderau, eds., *The YouTube Reader*. Stock-

holm: National Library of Sweden, 2009.

LOOKING FOR BACHELORS IN AMERICAN SILENT FILM: EXPERIMENTS WITH DIGITAL METHODS

Lisa Spiro

Whenever I test a new search interface, I look for "bachelor." I choose that word because my dissertation explored the relationship between bachelorhood and authorial identity in nineteenth century America, and the project still has a hold on me. When I used Lantern, the search tool for the Media History Digital Library (MHDL), to search for "bachelor" between 1855 and 1929 (up to the end of the silent film period), I found 5,282 results, which both overwhelmed and intrigued me.[1]

The sixth result drew my notice: "The diminutive form of the fairy on the table. The **bachelor**, although astonished, in reality sees nothing." Since I'm fascinated by the association between the bachelor and reverie, this result called for further exploration. It comes from a description of the trick photography used to create *Princess Nicotine; or, The Smoke Fairy* (1909), a Vitagraph short.[2] The film renders a character familiar from antebellum American sentimental works such as *Reveries of a Bachelor*: a gentleman (by cultural association a bachelor) dreaming over his pipe. In *Princess Nicotine*, two tiny, impish female fairies tease the bachelor. Eventually he torments a fairy in return, blowing smoke on her and jabbing toward her with his cigar. After she lights a fire, he sprays a water bottle wildly on the flames, the fairy, and ultimately himself. While antebellum bachelor reveries are typically sentimental, the female objects of this reverie playfully fight back, and its onanistic, comic nature can't be missed.

Watching this reimagining of the dreaming bachelor sparked my curiosity. What happens to the stereotype of the bachelor as an artist and dreamer in the early twentieth century, particularly in silent film? How might I make sense of more than five thousand

results and see, for instance, film genres and character types associated with "bachelor?" Thanks to Eric Hoyt and Derek Long, I got copies of text files of film magazines from 1905 to 1929 such as *Variety* (1905–26), *Moving Picture World* (1907–19), and *Exhibitors Herald* (1917–29): a total of 630 files or 1.97 GB.[3] In working with such a rich collection, I wanted to test digital methods such as n-gram analysis, topic modeling, and word frequencies. How might an "ordinary working" scholar, one without sophisticated programming skills, use digital methods to explore representations of bachelorhood in film magazines?[4] What are the pitfalls of digital methods, and what insights do they yield?

I aimed not to secure definitive answers, but to see how digital tools might fuel my own exploration across thousands of pages of text. As Ted Underwood suggests, we can "use text mining in an exploratory way, to map archives and reveal patterns that a critic could then interpret using nuanced close reading."[5] I didn't necessarily expect to find empirical evidence that there was a clear relationship between bachelors and artistic production, but instead to detect clues that would direct my inquiry.

In facing the challenge of scale, I turned to Joshua Sternfeld's recommendation that we should deal with digital abundance by drawing upon principles of appraisal, long important to archives.[6] He suggests that digital historical appraisal should consider scope (what is included in a collection) and provenance (where the material comes from and what its context is). Lantern clearly describes the scope of the MHDL, including available years, total pages, and circulation statistics for each title (if that information is available).[7] By using such information, a researcher can evaluate the significance of a term; does it appear more frequently in, say, the 1910s because there are more magazine volumes from that period in the MHDL?

I kept in mind two broad contexts in which my analysis operated:

the cultural and social position of US bachelors and the development of silent film. During the 1880s to early 1930s, a period Howard Chudacoff calls "The Age of the Bachelor," unmarried men flooded American cities, contributing to the rise of bachelor apartments to house them, saloons to entertain them, and restaurants to feed them.[8] Magazines, movies, and literature both included bachelors as characters and appealed to them as audiences, promoting "male consumerism."[9] Around the same time, film underwent significant developments as the Hollywood film industry emerged; genres such as dramas, westerns, and slapstick comedies became prominent; narrative techniques such as dynamic camera movement, close-ups, and dialogue intertitles were developed; feature films evolved; and the star system and fan magazines shaped careers.[10]

These contexts informed both the kinds of questions I asked and how I interpreted the results, but of course untangling the social, aesthetic, and cultural threads shaping the bachelor figure is complex. By experimenting with computational methods, I hoped to discern larger patterns across film magazines, isolating features that I could then interpret. My experiments with concordances, n-gram analysis, topic modeling, and text analysis suggest digital methods' potential for generating research questions, as well as the need for nuance in making any claims about results given the messiness of the data and the ambiguity of language.

EXPERIMENT 1: CREATING A CONCORDANCE TO UNDERSTAND THE SCOPE OF A CORPUS

To get a sense of how "bachelor" was used across my corpus, I used AntConc, a freeware package for creating concordances and conducting textual analysis.[11] A concordance lists words in a text or set of texts and typically shows the context around these words. Scholars have used concordances since the thirteenth century to discover patterns in texts, study word usage, and observe word frequencies.[12] Beginning in the late 1940s, humanities computing

pioneer Father Roberto Busa used computers to generate a lemmatized concordance of the works of Saint Thomas Aquinas and associated authors, listing all words under their dictionary headings.[13] Digital concording tools allow researchers to create concordances quickly and to ask questions iteratively.[14] AntConc offers several advantages over keyword search, including the ability to sort results and probe the contexts surrounding words. Creating a concordance for "bachelor" (ignoring case) across my corpus yielded 6,234 results, which I then sorted alphabetically to group related usages.

I discovered several factors that skew how frequently words appear, such as their repeated appearance in advertisements and in film and play listings. In glancing across the list of terms that occur with "bachelor," I noted optical character recognition (OCR) errors (such as "Bachelor'a") and recurring phrases (such as "bachelor dinner"). I quickly picked out titles of movies and plays, since "bachelor" is often in quotation marks and since the same phrase occurs multiple times (such as "The Bachelor's Baby"); also, the first letter (or all letters) of "Bachelor" tends to be capitalized.[15] For example, "The Bachelor's Baby" appears in files from 1909 to 1917, 1922, 1926, and 1927. Many of these titles are theatrical productions listed in *Variety* and the *New York Clipper*. Sometimes the same play (e.g., Francis Wilson's *The Bachelor's Baby)* was produced in multiple years (1910, 1911, 1912, and 1914). "The Bachelor's Baby" also was the title of films from 1913, 1915, 1922, and 1927 (including one British film), suggesting that film portrayals may have been influenced by vaudeville.

The concordance also called attention to the many ways that a term could be used. For example, variations of "bachelor apartment" occurred nearly three hundred times. Closer examination reveals references to a 1921 comic film with Georgia Hopkins, residences for men (including The Pageant, a St. Louis movie theater complex with bachelor housing), and a common film or vaudeville

setting where bachelors woo women (e.g., *The Great Question, 1915)* or live in comfort *(e.g., The Turn of the Wheel*, 1911). Film periodicals provide evidence not only for media history but also for social and cultural history, as we see how places associated with bachelorhood were advertised and imagined.

EXPERIMENT 2: EXPLORING BACHELOR TYPES USING N-GRAMS

While the concordance enabled me to glimpse how "bachelor" was deployed across my corpus, I wanted to examine more specific constructions of bachelorhood. I used AntConc's Clusters/N-gram feature to explore frequent two-word phrases where "bachelor" appears on the left or right of an adjacent word.[16] N-grams are broadly familiar thanks to the Google Books Ngram Viewer, which allows users to track the popularity of phrases across the Google Books corpus.[17] For example, Jean-Baptiste Michel et al. use n-grams to detect censorship, finding that the names of authors of philosophy, art, and politics books included on the Nazis' "degenerate" lists declined significantly in German works published from 1933 to 1945.[18]

On the left, "bachelor" typically functions as a modifier, such as "bachelor club." AntConc provides information about the "cluster" (n-gram), rank (how frequently the cluster appears in comparison to other n-grams), frequency (number of times the n-gram appears across the corpus), and range (number of files in which the n-gram occurs.) After removing n-grams in which stop words (common words like "the" that are filtered out of the analysis) appear, such as "bachelor and," I noticed that several of the top ten phrases seemed to describe people or places (see table 1). Ironic juxtapositions—the bachelor who is a father, girl, or bride—appear relatively often. The frequent appearance of bachelor clubs, apartments, and quarters seems to support Howard Chudacoff's claim that bachelor spaces were prominent in the early 1900s. Yet the common phrase "bachelor dinner" (a pre-wedding party) may

Rank	Frequency	Range	Cluster
2	303	77	bachelor club
3	301	36	bachelor daddy
4	219	106	bachelor dinner
6	182	67	bachelor apartments
9	92	31	bachelor girl
10	84	41	bachelor girls
14	51	14	bachelor brides
15	50	42	bachelor apartment
16	45	5	bachelor (com
19	40	31	bachelor quarters

Table 1. Bachelor n-grams, left. Phrases with stop words were omitted.

reflect an interest in the transition into marriage. We can also see a potential association between bachelor and comedy, as denoted by the abbreviation "com" that is used in listings of films such as *The Bachelor* (1911), *A Fascinating Bachelor* (1911), and *How to Catch a Bachelor* (1911). These n-grams raise a number of questions, such as why there were so many bachelor comedies in 1911 and why "bachelor" is attached to "daddy" and "girl."

But n-grams can be misleading. For example, the frequency of "bachelor daddy" seems to suggest a connection between bachelorhood and patriarchy. However, the phrase only occurs in 36 files, and almost all refer to a 1922 Alfred E. Green film. By examining exhibitors' notes and ads for *Bachelor Daddy*, I could study how films were promoted in the early 1920s, but I can't necessarily make broad claims about the association between bachelorhood and fatherhood.

Placing "bachelor" on the right side of an n-gram usually makes it the modified term, yielding a typology of bachelors as defined by their age, wealth, attitude toward marriage, or emotional status (see table 2). "Old bachelor," which dates back to at least 1630, means "an elderly or confirmed bachelor; *spec.* one having

Rank	Freq	Range	Cluster
3	173	80	old bachelor
4	149	15	small bachelor
7	137	76	young bachelor
8	114	67	wealthy bachelor
11	40	38	confirmed bachelor
13	33	26	rich bachelor
14	30	8	gay bachelor
15	27	13	lonely bachelor
16	27	4	reels. Bachelor
19	24	11	bashful bachelor

Table 2. Bachelor n-grams, right. Phrases with stop words were omitted.

the fastidious habits considered to be typical of such a person"; think Scrooge.[19] Magazines frequently label the "old bachelor" as "crusty," "wealthy," or "rich." A common plot shows a baby, girl, or young woman winning the heart of a crusty (and often wealthy) bachelor, resulting both in his domestication and her supposed salvation. For example, in *Hearts Asleep* (1919) a poor, noble orphan girl is rescued from criminals by a benevolent, wealthy bachelor who eventually marries her.

Whereas old bachelors are frequently depicted as crusty or remote, young bachelors are often shown as eligible men on the verge of marriage. Frequently they are wealthy, making them even more attractive. For example, in *Wanted, A Wife* (1912), a young bachelor must find a wife in order to keep his inheritance. Sometimes wealthy young bachelors are dissipated. For example, in *A Man and His Money* (1915), a "wealthy young bachelor" loses his fiancée after he wastes his fortune.

"Bachelor" is used as a code word for homosexual, but I did not find much clear evidence for this association in silent film descriptions, despite the frequency of "gay bachelor" and "confirmed

bachelor." Most of the results for "gay bachelor" refer to *A Gay Bachelor* (1911), a comedy about a man who woos a widow and then tries to get out of the marriage after discovering that she has five children.[20] "Gay" in my corpus typically implies "hedonistic,"[21] as with "a rather gay bachelor, given to clubs, parties and poker" (*Jane Marries*, 1913). An advertisement for *The Gay Bachelor* (aka *The Gay Lord Quex*, 1919) captures the association between "gay" and hedonism: "Midnight Parties/ Beautiful Women/ Gayest Action."[22] Whereas the gay bachelor is depicted as debauched, the "confirmed bachelor" is typically shown as resistant to marriage, whether because of crustiness (often in comedies) or early heartbreak (often in dramas). Typically, a film about a confirmed bachelor ends with him married (*Husbands Wanted*, 1911), the guardian of a child (*Little Girl*, 1911), or the victim of comic mishaps (*A Hazardous Courtship*, 1915).

However, by browsing uses of "bachelor," I stumbled across a few references in fan magazines from the late 1920s and early 1930s (edging beyond this study's time period) that draw an implicit connection between "bachelor" and "homosexual," even as they use a star's single status to attract fans. A 1930 gossip column announced "Don't go yet, girls, for I'm going to introduce you to somebody who has never found the Right Girl. But you will have to put on the best you've got and step fast, for Polly Moran is hanging around in the offing. . . . Stand up, William Haines, and let us see what a bachelor looks like."[23] While this description makes Haines, widely known in Hollywood as a gay actor, an object of pursuit, it may also send a subtle message about his sexuality. As Ronald Gregg argues, MGM publicists used codes to associate Haines with homosexuality in order to capitalize on "the public's growing fascination with homosexuality while also trying to protect themselves from the wrath of moral reformers."[24]

EXPERIMENT 3: EXPLORING CATEGORIES OF BACHELOR FILMS WITH TOPIC MODELING

While my experiments with concordances and n-gram analysis illuminated usage of "bachelor" across the corpus, I wanted to zoom into individual film synopses and use topic modeling to cluster them. Hence I constructed a Zotero collection of "bachelor films," drawing from synopses supplied in film reviews, exhibitors' guides, and occasionally advertisements. Since I didn't want to predetermine what counted as a bachelor (favoring, say, dreaming bachelors over lecherous ones), I applied broad criteria—any film synopsis or review in which a bachelor appeared as a character. I used a slow, manual process to extract film descriptions, subject to human error and inconsistencies and constrained by the idiosyncrasies of digitized magazines.[25] Sometimes film descriptions break over pages or are interrupted by advertisements or images, making it difficult to extract a coherent chunk of text. I also ran into a number of OCR errors in some synopses; if there were so many that the text wasn't comprehensible, I skipped it.

By searching for "bachelor" across my corpus, I found 774 bachelor films: 26 from 1906 to 1909, 640 from 1910 to 1919, and 108 from 1920 to 1929.[26] While this list seems to point to the ubiquity of the bachelor figure, I readily acknowledge its flaws. My corpus, which begins in 1903, contains more magazines from the 1910s than any other period, so films from this decade are overrepresented. Undoubtedly my filmography includes many false positives, films in which a bachelor is just an incidental character.[27] In addition, I may be missing some bachelor films, as the term "bachelor" may not appear in *all* magazine descriptions of such films, or it may be obscured through OCR errors. For example, the term "bachelor" is not used in any magazine descriptions of *Princess Nicotine* in my corpus, just in a 1912 book in the MHDL.

To explore what categories might emerge from my film synopses, I used probabilistic topic modeling. Latent Dirichlet allocation (LDA), a popular topic modeling method, uses statistical techniques to identify terms that frequently appear together (topics)

and calculates what percentage of each topic is included in a particular document.[28] Eric Hoyt employs topic modeling to explore film magazines in the MHDL corpus, discovering content that might otherwise escape the notice of a critic. For example, topic models reveal that *Variety* published a number of advertisements for rental housing.[29] By grouping together related words into topics, topic modeling can reveal significant patterns. However, as Benjamin Schmidt points out, topic modeling can lead to misleading results because clusters don't necessarily cohere beyond the first ten words or so. Further, topics often aren't stable across time; not only does the meaning of words change, but so do the words documents use.[30]

Taking these cautions into account, I tested topic modeling with my filmography. For ease, I used Topic Modeling Tool, a graphical user interface for the topic modeling package Mallet.[31] While Topic Modeling Tool was easy to use, I was aware of several factors that might distort my results, including the variable length and content of synopses (which may have been too short to yield meaningful results), OCR errors, partial synopses (such as those interrupted by page breaks), and synopses corrupted by chunks of extraneous texts such as ads. The messiness of the data is evident in a few of the resulting topics, where OCR errors like "tbe" (for "the") or partial words like "ing" rank among the top ten words. I chose not to correct this messy data, partly because of time constraints, partly because I wanted to explore the challenges of working with such data. However, I did remove 16 files that were 12 kilobytes or more from my analysis, since Topic Modeling Tool ranks documents based on the *number* of words linked to a topic and gives greater emphasis to longer documents. This left 758 movie synopses ranging from 118 bytes (around 118 characters) to 7 kilobytes (approximately 7,000 characters). I created 20 topic models, as fewer would not allow me to make fine-grained enough comparisons, while more would be unwieldy. With each topic, I reviewed the text of the top four results to discern com-

mon elements. No film synopsis belonged exclusively to one topic; a synopsis in the top four results contained between 11% and 50% of the topic.

I naively expected all films to be classified by genre (such as comedy or melodrama), but no clean, coherent categories emerged. However, many topics seemed to *emphasize* elements such as genre, setting, character, plot, and/or film language. For example, "bachelor maid de bride pretty police tom hands head fire" seems to be associated with the genre of melodrama, as words like "police" and "head" may indicate. Likewise, "made married son comedy robert fun amusing attention frank" describes comic films, as words like "amusing," "comedy," and "fun" suggest. "Widow girls country heart day boys doctor marry life tom" seems to reflect films *set* in the country, which is often depicted as a more virtuous place than the city. Some topics seem to focus on types of *characters*. For example, an innocent, independent woman often plays a prominent role in films in the topic "woman young lady life man sister years point make world." Others are loosely related by *plot*. For example, "love bachelor mr falls ruth wealthy party lost good happy" appears to be associated with films about bachelors falling in love. Since many film summaries offer criticism as well as plot summaries, a few topics are particularly imbued with the language of *film reviews*. Witness, for example, "story picture good part miss star cast production screen role."

Given the heterogeneity of the film descriptions, which are of varying (generally short) lengths and from different periodicals and years, topic modeling did not produce much clarity about types of bachelor films, beyond showing the diversity of such works. However, it did suggest potential research topics, including investigating the emergence of film criticism, the relationship among setting, character, and genre, and the significance of women in bachelor films. I was reminded that film synopses contain words that reflect a range of elements, including what

happens, where it happens, who is involved, and how reviewers received the film.

EXPERIMENT 4: CONTEXTUALIZING WORDS USING TEXT ANALYSIS

In examining these topics, I wondered how the words constituting them were used. I used the text analysis tools Voyant and AntConc to explore the most frequently occurring words in my bachelor filmography.[32] Through Voyant, I quickly saw that a few of the most common words in my filmography were missing from my topics: "old," "little," and "comes." All three are considered to be Mallet stop words and are automatically ignored by Topic Modeling Tool unless the settings are adjusted.[33] Yet these words are significant. For example, "old" is the fifth most frequent word when TAPoR stop words are used, occurring 540 times. Not only does "old" frequently refer to old bachelor(s) (94 instances), but also to "old man" (52) and "old maid" (45).

Hence the topic models discussed in the previous section give a distorted view of the corpus, de-emphasizing film descriptions in which words such as "old" and "little" appear. To see what difference employing a more restricted stop word list would make, I generated new topic models using the Taporware list,[34] which is shorter and contains fewer content words like "old." Not surprisingly, the topics were significantly different—in part because probabilistic topic modeling is dynamic, generating new results every time the program is run, but also because a more expansive vocabulary was included in the analysis. For example, the topic "old love widow young letter bis maid bachelor arms mother" seems to describe films about bachelors wooing widows and other women. As with the previous set of topics, the clearest topics generated with the Taporware list reflect the language of film criticism, such as "story picture good play cast role production screen scenes interest." Comparing the topics generated with the Taporware and Mallet stop word lists illustrates the fuzziness

of topic modeling. In only 15% of cases did the topics on both lists contain at least two words in common and share at least two films that were included in the top four results. For example, "daughter marry make love billy pretty rich money taken place" (Taporware) resembles "home children billy money death marie order position society bob" (Mallet stop words); both topics seem to be about temptation and sin, and both include *Jackstraws* and *Dangerous to Men* in the top four results. Comparing results from different stop word lists demonstrates the importance of looking inside the black box and understanding how text analysis tools work. With short, messy, and diverse texts like these film synopses, topic modeling may point to features such as the presence of critical or sensational language, but it often does not produce clear classifications.

While topic modeling clusters seemingly related texts, examining keywords in context reveals the range of ways that a word is used across texts. Take, for example, the Mallet stop words topic I identified as "melodrama," which includes words like "bride," "police," "head," and "hands." This chain of words made me imagine a panicked bride with her hands on her head who is rescued by police. Yet many words carry multiple connotations. For example, "hands" can function as a noun or a verb, reflecting how the body is associated with emotion, appearance, and action. Shaking hands signify fear, while clasped hands indicate friendship and even marriage. As an idiom, "hands" indicates responsibility or control, such as a girl in the hands of criminals. "Hands" can also describe an action, as when a ring is handed back or a letter is handed to its recipients. In westerns, "hands" are workers: ranch hands. Not all of these meanings are necessarily captured by the melodrama topic, since topic modeling takes into account context by grouping words that co-occur. However, such an example points to the need to examine context. It also suggests a research question investigating the language of the body in silent film descriptions.

CONCLUSION: REFLECTIONS ON BACHELOR MACHINES

The bachelor seems to be an engine of plot, driving comic scenarios in which he becomes a baby's guardian, sentimental ones in which his crusty heart is melted by a young woman, and sensational ones in which he lures a woman to his bachelor apartments for tea (and more). In 1928, film critic and director Norbert Lusk acerbically commented on the ubiquity of bachelor scenarios and their underlying creepiness:

> Two rather musty — no, very musty — situations inspire the picture known as 'Beau Broadway.' . . . When a worldly bachelor is asked to look after the granddaughter of a dying friend, promptly assumes that she is a child and discovers her to be an ingénue with lots of sex appeal, I recall two hundred and forty-six versions of this in novels, plays, and movies. And when the roué marries the ingénue in the end, I find the conjunction repellant.[35]

The romance of a roué and an ingénue seems to combine comedy (as the bachelor discovers the baby is actually a beauty), titillation, and marriage. Other common bachelor plots include bachelors (sometimes in threes) adopting babies and bachelors pretending to be married in order to win an inheritance. As a scathing review of *Oh Baby* (1926) noted, "Of all the over-worked themes for cinema farce, the most constantly used is probably that of the young bachelor who must pose as a married man in order to win a fortune from a rich old relative."[36] Indeed, many comedies generate laughter through bachelors scrambling to impersonate married men or fathers—and often end in them becoming what they once resisted.

As the Hollywood machine churned out new plots from elements of old ones, the bachelor figure helped fuel it. "Frisky bachelor compromises housemaid chivalry betrothal."[37] Such a plot was supposedly constructed using the Movie Writer, "a scenario for-

ming device" patented in 1915 by Arthur F. Blanchard. The Movie Writer both entertained and stimulated ideas for films and other creative works by putting terms into relation with each other.[38] Using rollers, the machine feeds through six strips, each one printed with (in succession) an adjective, noun, verb, noun, denouement word, and conclusion word. By turning the spindles, users randomly selected terms to construct a plot description. A *Boston Post* article cited by Epes Winthrop Sargent describes a scenario that a "confirmed bachelor" supposedly created using the Movie Writer: 'Eccentric - Spinster - Adopts - Burglar - Excitement – Marriage.'[39] The absurdity of this plot exposes how commonly bachelors and spinsters appeared in film; bachelors sometimes either were burglars or fought them off, and they frequently adopted girls and ended up married to them.

We might regard the Movie Writer as a bachelor machine, which produces film plots mechanistically and reductively rather than through human creativity and agency.[40] As Constance Penley argues, the bachelor machine often involves "the dream of the mechanical reproduction of art" and is "typically a closed and self-sufficient system."[41] But Blanchard insisted that the Movie Writer's purpose was to spark rather than supplant creativity, and Sargent likewise agreed that this "thinking game" could produce interesting ideas through "accidental combination."[42] On the one hand, the Movie Writer reduces film scenarios to a few artificially generated elements, but on the other it could foster new ideas from common terms.

Like the Movie Writer, computational analysis of film magazines can be both reductive (revealing patterns without context) and generative (prompting further exploration and interpretation). Each observation led to me to new research questions, such as about the connection between bachelor characters in vaudeville and silent film, the marketing of film, the elements of genre, and the language of the body in film reviews. By using text analysis

tools, I paid attention to varieties of bachelorhood that I might have otherwise ignored, such as young bachelors. Delving into the MHDL led to an immersive, if incomplete, education in film history for someone whose background is literature. I observed the connections between vaudeville and silent film, the importance of genres such as westerns, melodramas, and comedies, and the language of film criticism.

Yet I am also more aware of the tenuousness of any claim I might make based on using text analysis with a collection as diverse as the MHDL. I see how messy data can be, how much interpretation depends on context, and how the length and nature of the texts determine results. As Sternfeld reminds us, scope matters: the relative paucity of 1920s films affected my results, as did the diversity of content and the shorter length of many descriptions. I also saw the importance of context. A word might appear frequently because it is used repeatedly in exhibitors' notes and advertisements. I appreciate the need to move between the macro and micro view, to observe larger patterns and then dive into texts to make sense of them. I also saw the limitations of the off-the-shelf tools that I used, which didn't provide a straightforward way for me to study changes in representations of bachelors over time. I understand how time consuming it can be to get data into the proper form (such as by extracting movie summaries from text versions of magazines) and analyze that data (such as by trying to make sense of topic modeling results).

This project is preliminary. It demands a more substantial grounding in film, cultural, and social history and theory, as well as more sophisticated methods of text analysis. My initial experiments have demonstrated the challenges of working with a messy but rich corpus and crude but illuminating methods, as well as the possibilities opened up by going beyond keyword search. Based on the evidence uncovered in the MHDL, I believe that bachelors are a significant character type in silent film, although I found

fewer bachelor dreamers and artists than I expected. I want to further explore how depictions of the bachelor changed between the late nineteenth century and 1929. Staring into a flickering screen filled with text analysis results can produce its own kind of reverie, less fantasy than a sense of curiosity and potential.

ENDNOTES

1 This essay is based on my presentation at the 2015 Arclight symposium. Thanks to Charles R. Acland and Eric Hoyt for inviting me to participate in the symposium and for providing helpful feedback on this essay, as well as to my fellow participants for stimulating presentations and conversations.

2 Frederick Arthur Ambrose Talbot, *Moving Pictures: How They Are Made and Worked* (J.B. Lippincott, 1912)*; Princess Nicotine* (Vitagraph, 1909).

3 For a list of files included in this study, as well as other supplemental data associated with this essay, please see http://hdl.handle.net/1911/81720.

4 Dan Cohen et al., "Data Mining with Criminal Intent," 2011.

5 Ted Underwood, "For Most Literary Scholars, Text Mining Is Going to Be an Exploratory Tool," *The Stone and the Shell,* accessed April 7, 2015, http://tedunderwood.com/2011/08/15/how-to-make-text-mining-serve-literary-history-and-not-the-other-way-around/.

6 Joshua Sternfeld, "Historical Understanding in the Quantum Age," *Journal of Digital Humanities,* August 25, 2014, http://journalofdigitalhumanities.org/3-2/historical-understanding-in-the-quantum-age/.

7 See http://lantern.mediahist.org/

8 Howard P. Chudacoff, *The Age of the Bachelor: Creating an American Subculture* (Princeton University Press, 2000).

9 Ibid., 216.

10 William K. Everson, *American Silent Film* (Da Capo Press, 2009); Eileen Bowser, *The Transformation of Cinema, 1907–1915* (University of California Press, 1994).

11 Laurence Anthony. "AntConc." Accessed August 9, 2015. http://www.laurenceanthony.net/software/antconc/.

12 Geoffrey Rockwell, "What Is Text Analysis, Really?," *Literary and Linguistic Computing* 18, no. 2 (June 1, 2003): 209–19, doi:10.1093/llc/18.2.209.

13 Susan Hockey, "The History of Humanities Computing," in *A Companion to Digital Humanities*, ed. Ray Siemens, John Unsworth, and Susan Schreibman, Blackwell Companions to Literature and Culture (Oxford: Blackwell, 2004), http://www.digitalhumanities.org/companion/.

14 Rockwell, "What Is Text Analysis, Really?"

15 Typically, text mining involves a preprocessing step where punctuation and capitalization are removed, but for my purposes this information is important to detecting titles and other proper nouns.

16 I ignored case in creating the n-grams so that "Bachelor Daddy" and "bachelor daddy" would be combined into the same result. However, spacing affected the results; for example, I got separate results for "old bachelor" and "old bachelor" (with one more space between the terms). Additional cleaning of the corpus is necessary.

17 "Google Ngram Viewer," http://ngrams.googlelabs.com/.

18 Jean-Baptiste Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books," *Science* 331, no. 6014 (January 14, 2011): 176–82, doi:10.1126/science.1199644.

19 "Old, Adj.," *OED Online* (Oxford University Press), accessed August 31, 2015, http://www.oed.com/view/Entry/130955.

20 "A Gay Bachelor," *The Moving Picture World*, September 9, 1911, 717, MHDL.

21 "Gay, Adj., Adv., and N.," *OED Online (Oxford University Press)*, accessed August 15, 2015, http://www.oed.com/view/Entry/77207.

22 "J.H. Kunsky's Detroit Theatres Advertised in Distinctive Style," *Exhibitors Herald*, January 24, 1920, MHDL.

23 Homer Croy, "We Have with Us Tonight," *The New Movie Magazine*, February 1930, 115, MHDL.

24 Ronald Gregg, "Gay Culture, Studio Publicity, and the Management of Star Discourse: The Homosexualization of William Haines in Pre-Code Hollywood," *Quarterly Review of Film and Video* 20, no. 2 (January 2003): 88. doi:10.1080/10509200390118489.

25 In working with text files, it wasn't always clear where synopses

come from; looking at page images helps to establish context, but slows moving through results. I was somewhat inconsistent in how I extracted descriptions; while sometimes I took the complete text, sometimes I omitted information such as the cast list.

26 This corpus lacked several magazine titles that were part of my final corpus, including the *New York Clipper, Film Daily,* and *Motion Picture News* from 1917 to 1929. If a description occurred in multiple magazines, I generally took the first one I encountered. I accidentally included two synopses for *Bachelor Bill's Birthday Present* (1913), which may have slightly skewed results.

27 While I tried to focus on American films, my filmography includes at least a few films from other countries, including the UK.

28 David M. Blei, "Probabilistic topic models," *Communications of the ACM* 55, no. 4 (April 1, 2012): 77. doi:10.1145/2133806.2133826.

29 Eric Hoyt, "Lenses for Lantern: Data Mining, Visualization, and Excavating Film History's Neglected Sources," *Film History* 26, no. 2 (2014): 146–68.

30 Benjamin M. Schmidt, "Words Alone: Dismantling Topic Models in the Humanities," *Journal of Digital Humanities,* April 5, 2013, http://journalofdigitalhumanities.org/2-1/words-alone-by-benjamin-m-schmidt/.

31 David Newman and Arun Balagopalan, *Topic Modeling Tool,* https://code.google.com/p/topic-modeling-tool/.

32 Stéfan Sinclair, Geoffrey Rockwell, and the Voyant Tools Team. *Voyant Tools* (web application), 2012, http://voyant-tools.org/.

33 David Mimno, "Multilingual Stop Lists," *GitHub,* accessed August 31, 2015, https://github.com/mimno/Mallet.

34 Taporware Stop Words, https://github.com/sgsinclair/trombone/blob/0878df1a2a80cc64c7b164500ca528b0600e911e/src/main/resources/org/voyanttools/trombone/keywords/stop.en.taporware.txt.

35 Nobert Lusk, "The Screen in Review," *Picture-Play Magazine,* November 1928, 99, MHDL.

36 "Newspaper Opinions," *The Film Daily,* August 16, 1926: 8, MHDL.

37 "A Thinking Machine," *The Literary Digest,* July 22, 1916, Google Books.

38 Arthur Franklin Blanchard, Movie-Writer, US Patent 1198401, filed

October 23, 1915, and issued September 19, 1916, accessed May 17, 2015, http://www.freepatentsonline.com/1198401.html.

39 Epes Winthrop Sargent, "The Photoplaywright," *Moving Picture World*, November 20, 1915: 1492, MHDL.

40 Thanks to Mark Williams for his suggestions about the bachelor machine.

41 Constance Penley, *The Future of an Illusion: Film, Feminism, and Psychoanalysis*, *Media & Society* (Minneapolis: University of Minnesota Press, [1989], 1989), 57.

42 Sargent, "The Photoplaywright," 1492.

BIBLIOGRAPHY

"A Gay Bachelor." *The Moving Picture World*, September 9, 1911, 717. MHDL.

"A Thinking Machine." *The Literary Digest*, July 22, 1916, 182–83. Google Books.

Blanchard, Arthur Franklin. Movie-Writer. US Patent 1198401, filed October 23, 1915, and issued September 19, 1916. Accessed May 17, 2015. http://www.freepatentsonline.com/1198401.html.

Blei, David M. "Probabilistic Topic Models." *Communications of the ACM* 55, no. 4 (April 1, 2012): 77. doi:10.1145/2133806.2133826.

Bowser, Eileen. *The Transformation of Cinema, 1907–1915*. Berkeley: University of California Press, 1994.

Chudacoff, Howard P. *The Age of the Bachelor: Creating an American Subculture*. Princeton: Princeton University Press, 2000.

Cohen, Dan, Fred Gibbs, Tim Hitchcock, Geoffrey Rockwell, Jörg Sander, Robert Shoemaker, Stéfan Sinclair, et al. "Data Mining with Criminal Intent." Final white paper, 2011.

Croy, Homer. "We Have with Us Tonight." *The New Movie Magazine*, February 1930: 43, 115. MHDL.

Everson, William K. *American Silent Film*. Boston: Da Capo Press, 2009.

"Gay, Adj., Adv., and N." *OED Online*. Oxford University Press. Accessed August 15, 2015. http://www.oed.com/view/Entry/77207.

Goldstone, Andrew, and Ted Underwood. "The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us."

New Literary History 45, no. 3 (2014): 359–84.

"Google Ngram Viewer." Accessed May 21, 2011. http://ngrams.googlelabs.com/.

Gregg, Ronald. "Gay Culture, Studio Publicity, and the Management of Star Discourse: The Homosexualization of William Haines in Pre-Code Hollywood." *Quarterly Review of Film and Video* 20, no. 2 (January 2003): 81–97. doi:10.1080/10509200390118489.

Hockey, Susan. "The History of Humanities Computing." In *A Companion to Digital Humanities*, edited by Ray Siemens, John Unsworth, and Susan Schreibman. Oxford: Blackwell, 2004. http://www.digitalhumanities.org/companion/.

Hoyt, Eric. "Lenses for Lantern: Data Mining, Visualization, and Excavating Film History's Neglected Sources." *Film History: An International Journal* 26, no. 2 (2014): 146–68.

"J.H. Kunsky's Detroit Theatres Advertised in Distinctive Style." *Exhibitors Herald*, January 24, 1920. MHDL.

Laurence Anthony. "AntConc." Accessed August 9, 2015. http://www.laurenceanthony.net/software/antconc/.

Lusk, Nobert. "The Screen in Review." *Picture-Play Magazine*, November 1928, 68–71, 98–99. MHDL.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, et al. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331, no. 6014 (January 14, 2011): 176–82. doi:10.1126/science.1199644.

Mimno, David. "Multilingual Stop Lists." *GitHub*. Accessed August 31, 2015. https://github.com/mimno/Mallet.

Newman, David and Arun Balagopalan. *Topic Modeling Tool*. https://code.google.com/p/topic-modeling-tool/.

"Newspaper Opinions." *The Film Daily*, August 16, 1926, 8. MHDL.

"Old, Adj." *OED Online*. Oxford University Press. Accessed August 31, 2015. http://www.oed.com/view/Entry/130955.

Penley, Constance. *The Future of an Illusion: Film, Feminism, and Psychoanalysis*. Media & Society. Minneapolis: University of Minnesota Press, 1989.

Princess Nicotine. Vitagraph, 1909.

Rockwell, Geoffrey. "What Is Text Analysis, Really?" *Literary and Linguistic Computing* 18, no. 2 (June 1, 2003): 209–19. doi:10.1093/llc/18.2.209.

Sargent, Epes Winthrop. "The Photoplaywright." *Moving Picture World*, November 20, 1915. MHDL.

Schmidt, Benjamin M. "Words Alone: Dismantling Topic Models in the Humanities." *Journal of Digital Humanities*, April 5, 2013. http://journalofdigitalhumanities.org/2-1/words-alone-by-benjamin-m-schmidt/.

Sinclair, Stéfan, Geoffrey Rockwell and the Voyant Tools Team. *Voyant Tools* (web application). 2012. http://voyant-tools.org/.

Sternfeld, Joshua. "Historical Understanding in the Quantum Age." *Journal of Digital Humanities*, August 25, 2014. http://journalofdigitalhumanities.org/3-2/historical-understanding-in-the-quantum-age/.

Talbot, Frederick Arthur Ambrose. *Moving Pictures: How They Are Made and Worked*. J.B. Lippincott, 1912. MHDL.

Underwood, Ted. "For Most Literary Scholars, Text Mining Is Going to Be an Exploratory Tool." *The Stone and the Shell*. Accessed April 7, 2015. http://tedunderwood.com/2011/08/15/how-to-make-text-mining-serve-literary-history-and-not-the-other-way-around/

TERMINOLOGICAL TRAFFIC IN THE MOVIE BUSINESS

Charles R. Acland and Fenwick McKelvey

A foundational commitment of Project Arclight has been the consideration of scholarly questions, methods, and curiosities at every stage of development. We understood, from the outset, that the design and usefulness of our digital tools had to be guided by the demands of media historians. Not only did our initial project design reflect this, but several participants tested and experimented with our search app as we progressed. The results helped us make further improvements at the level of software design and data visualization. And our tests demonstrated some of the most likely applications, benefits, and limits of our research tool.

In one such pilot study, we examined how, through use of the Arclight app, the Media History Digital Library (MHDL) corpus might provide insights into the entertainment industry's business, investment, and promotion terminology.[1] One of the features of our age is that there is considerable traffic between industrial and popular audiences about the entertainment business. We take for granted that success now comes with discussion of box-office revenues, budgets, and profits.[2] The expansion of this shared language speaks to the high degree of reflexive understanding of the capitalist core of the entertainment industry. This mutual vocabulary captures the discursive mechanisms in play with which "the business" is addressed and understood as such not only by industrial agents but also by fans, patrons, and broad publics.

Acland's recent work on the concept of the blockbuster shows that the term references more than a particular kind of hybrid genre of popular moviemaking; it involves and condenses ideas about budgets, financial success, merchandise, technology, and investment. In fact, so expansive is the term, and so connotatively powerful in our current phases of industrial cultural production, that

it is better to speak of a "blockbuster strategy" whose core characteristics are grasped by agents industrial and popular alike.[3]

Through genealogical work on this topic, we know that before 1942 there was no such thing as a blockbuster cultural product, named as such, though there were many lavish and expensive Hollywood productions. Given that "blockbuster" was a term that came to be shared by trade, promotional, and popular domains, what were some—if any—comparable terms prior to the initial circulation of that term? Using the Arclight search app, what sort of shared language of business, economies of entertainment, or financial success can we identify in the years before blockbuster moves into its lasting position as a bridge between industry and audiences?

In searching the MHDL corpus, we chose our dates not by accident nor as a default of the original use of blockbuster in 1942. The first uses of the term were not even in the realm of entertainment but in news reporting of military activity in Europe. Instead, looking at the existing corpus of the MHDL, we could see that there was a period of overlap between industry and fan publications for the years 1934 to 1941 with an acceptable level of overlap for 1931 to 1933. The digitization is extensive for these publications for those periods, so we understood that they offered a better basis for comparison between industrial and fan work than other date ranges in which the overlap of available digitized material was spottier. Keeping an eye on the best overlap for these years, we chose to concentrate on the following industry publications: *Film Daily, Independent Exhibitors Film Bulletin, Motion Picture Daily, Motion Picture Herald, Showmen's Trade Review,* and *Variety*. The fan publications selected were *Hollywood, Modern Screen, Motion Picture, New Movie Magazine, Photoplay, Picture Play, Radio Mirror,* and *Screenland*. Apart from *Film Daily* and *Motion Picture Daily*, all the trade periodicals were published on a weekly basis; the fan publications were all released monthly.

We now needed a list of terms. Led by graduate research assistant Tyler Morgenstern, we systematically sampled these publications through those years, collecting any term that specifically referenced the business or promotional features of the entertainment industry, regardless of the medium, though this corpus tilts considerably toward the movies. The fan magazines had considerably few such terms—no surprise there—with about 87 terms identified; there were 232 from the trades (see tables 1 and 2).

Some terms have multiple uses and connotations, and as such they would not give us a sense of the specific presence of a discourse about "show business." Disambiguation was carefully assessed, and terms whose definitions were not as specific to our concerns as possible were not further examined. Our preliminary study sought to understand the migration of terms between fan publications and industry publications. Since we conducted the study during the development of the Arclight tool, we selected a sample of keywords because the entire list would have overwhelmed the tool at the time. (Readers no longer have to worry about these early problems). For the sake of our pilot test, then, we selected seven terms to run through the Arclight search app, terms that we felt, with lesser degrees of ambiguity, signaled industrial operations: *box office, contract, flop, hit, profit, budget,* and *gross*. The first five appeared on both lists and were searched in their multiple spelling (e.g., *box-office* as well as *box office*) but the other two were limited to the industrial trade list.

Absorption	Agreement	Average gross
Accountants	Amortization	Back salary
Accruals	Anti-trust	Backers
Admission	Anticipated revenue	Balance sheets
Advantage	Arbitration	Bankroll
Advertise	Arrears	Banner
Advertising	Assembly line	Bargain
Affairs	Assets	Bargaining
Affiliates	Attraction	Best seller
Agency	Audit	Best selling

Bidder
Bids
Big budget
Biggest opening
Billings
Block booking
Bonds
Bonus
Bonuses
Bookings
Borrow
Box office
Box-office
Breach of contract
Budget
Business proposition
Businesses
Buyers
Capital stock
Capitalize
Grosses
Hired
Hit
Holdings
Holdover
Hookups
Hour-wage
Import
Cash in
Champions
Charges
Cheap
Cheaper
Clientele
Closed market
Coffers
Commission
Commodity
Compensation
Competition
Competitive advantage
Efficiency
Employment
Endorsements
Estimate
Exchange
Executives
Expenditure
Expenses
Expensive
Exploit
Exportation
Fee
Financial
Financier
Firm
First class
Fixed assets
Flop
Floppo
Flops
Foreign grosses
Foreign sales
Fractional
Freelancing
Gain
General strike
Gross
Grosser
Importations
Income
Incorporation
Increase
Indebtedness
Independent
Industrials
Industry
Interests
Invest
Labor
Layoffs
Lease
Liabilities
License fee
Liquid condition
Liquidation
Liquidity
Losses
Manager
Manufacturer
Market
Mass production
Merchant
Merger
Money raising
Monopolistic
Monopoly
Mortgage
Nationalization
Negotiated
Net
Opening
Operating profit
Opportunity
Outlay
Overhead
Overpayment
Overproduction
Overtime
Owners
Parent firms
Patronizing
Patrons
Payable
Payoffs
Payroll
Per cent
Percentage
Pirated
Stalemate
Stimulating
Stock certificates
Stock market
Stockholder(s)
Strike
Studio
Stunt
Subscription
Subsidiaries Subsidiary
Subsidized
Successful
Surplus

Takeover	Purchase	Shares
Takings	Purchaser	Shutdowns
Tariff	Purchasing	Slump
Tax	Rave	Sold
Taxation	Receipts	Speculation
Tie-in	Receivership	Sponsored
Trade	Record	Unemployment
Transfer	Renewals	Union
Treasury	Rentals	Valuable
Tremendous	Repayment	Value
Trust	Resources	Venture
Turnout	Returns	Wage
Prestige	Revenue	Wage sales
Prices	Royalty	Wage-employment
Private sale	Salaried	Wage-hour
Probabilities	Salaries	Waste
Proceeds	Sales	Wealth
Product	Salesman	Wholesale
Profit sharing	Saving	Wholly owned subsidiary
Profitably	Scheme	Win
Profits	Self-sustaining	Winner
Properties	Selling	Working capital
Publicity	Sensation	
Publishing	Settlement	

Table 1. Financialization key terms: industry publications. Drawn from *Film Daily, Independent Exhibitors Film Bulletin, Motion Picture Daily, Motion Picture Herald, Showmen's Trade Review, Variety* (1931–41).

Advertisers	Crash	Fortune
Assets	Debt	Free-lancing
Ballyhooed	Distribution	Gamble
Bid	Dollar figures	Grossed
Bonds	Earnings	Hits
Boom days	Economize	Income
Box office	Efficiency	Installment
Box office reports	Executives	Interest
Box office success	Expenses	Investment
Box-office	Failure	Job
Collective bargainer	Financial worries	Lavish
Contract	Financially	Loaned
Contract offer	Fiscal year	Longest run
Cost	Flop	Lucrative
Costly	Foreign sales	Lump sum

Manufacturer	Properties	Sensation
Market	Publicity	Signing
Money	Publicized	Smash
New market	Receipts	Sponsored
Offer	Record-breaker	Stock
On loan	Records	Stockholders
Opportunities	Returns	Success
Over the top	Rhapsodies	Taxed
Over-selling	Roster	Tendered
Paid	Salary	The industry
Payment	Salaries	Triumphant
Popularity	Screen product	Unsuccessful
Production plans	Screen rights	Wall Street
Profit	Sell	Work

Table 2. Financialization key terms: fan publications. Drawn from *Hollywood, Modern Screen, Motion Picture, New Movie Magazine, Photoplay, Picture Play, Radio Mirror, Screenland* (1931–41).

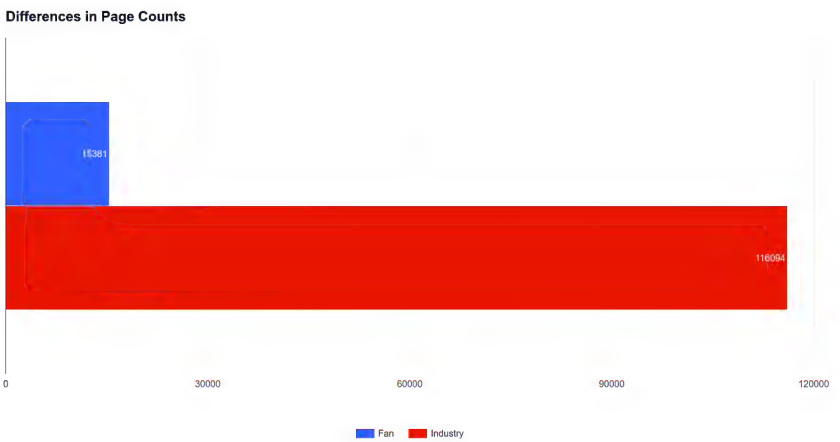


Figure 1. Page counts for selected keywords using Arclight app.

We ran each term through Arclight to get raw numbers of page counts by year and by publication. Arclight uses page count as its metric of relevance of each entity or keyword. In other words, it counts the number of pages mentioning a term rather than the overall frequency of the term in a publication.[4] Figure 1 charts the returns for all our selected keywords for industry and fan publica-

tions. Industry publications were clearly more prolific, no doubt in part due to the fact that they were published more frequently than the monthly fan magazines. Industry publications mentioned our terms on 116,094 pages from 1931 to 1941. The Arclight app returned 15,381 pages in fan publications that included terms from our sample. We should note that we could not normalize the results at that stage in the development of the Arclight app so we do not include any charts comparing activity between fan and industry publications. We hope in further studies to be able to contextualize (or normalize) the results within the overall monthly page counts of the publications, but the results here already demonstrate the potential of Arclight to understand macro-trends in film history.

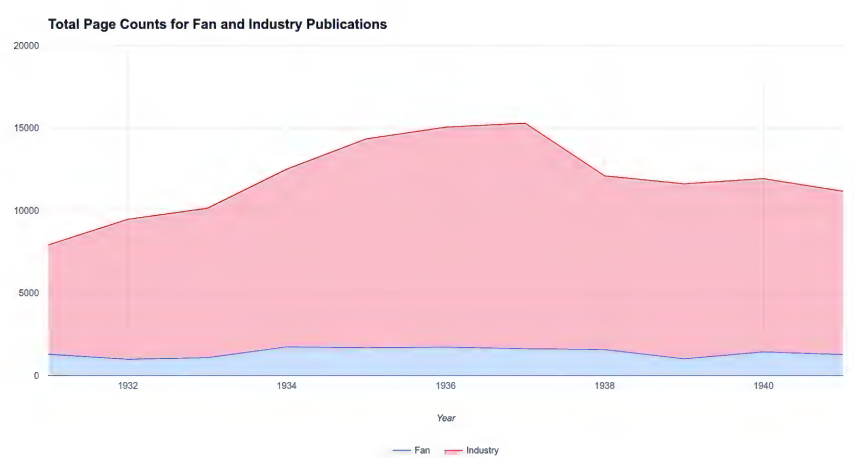


Figure 2. Total page counts for fan and industry publications.

Our early results revealed the changing language of film success. Figure 2 depicts the change in page counts per sector from 1931 to 1941. Industry publications gradually increase using terms in our sample, reaching a peak in 1937. The drop is not the result of a gap in the corpus as records exist for all publications, so other factors might be at work. The chart doesn't give an explanation,

but it does suggest the years 1935 to 1938 require more attention. Conversely, fan publications used our specified keywords uniformly through our period of study. We can use Arclight to further explore trends in greater detail. Figures 3 and 4 track our terms from 1931 to 1941 for both the industry and fan publications. Each line plots the changes per year in the page counts of a term.

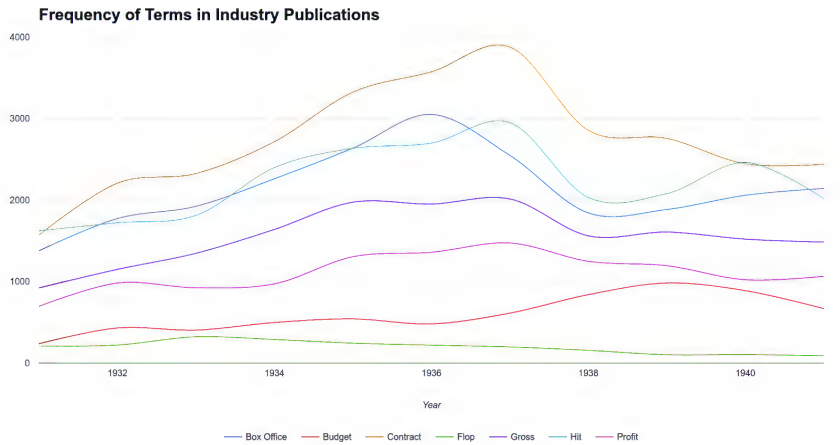


Figure 3. Term appearances in industry publications, 1931–41.

What can we learn from these charts and these research procedures? Failure, it seems, was an unpopular subject. Industry publications either avoided flops or avoided talking about flops: it was the only keyword that declined in use from 209 mentions in 1931 to just 88 mentions in 1941. Future studies might compare *flop* to other terms for failure (we know many) in order to see if failure overall was an unpopular subject or if *flop* simply flopped. Industry publications preferred *hit* (the second most frequent keyword) as well as *box office* (third most popular). *Profit*, ever a constant, also increased slightly in use, even after the drop in 1937. *Budget* is the only term that goes against the trend and actually increases in use after 1937. While not an explanation, the chart helps us pose a number of questions related to the discussion of finance in industry publications. Is there a relationship between *profit* and

box office given their correlation? Why did *profit*—the term, that is—increase beginning in 1936?

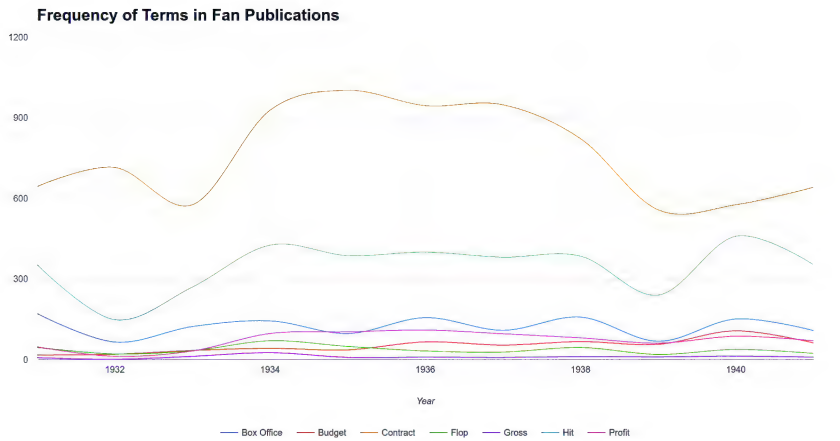


Figure 4. Term appearances in fan publications, 1931–41.

Fans seemed to enjoy hearing about contracts, hits, and little else. Figure 4 graphs all the terms in the fan publications, and it's important to note that we observe little migration of any financial terms. The data does not support our hypothesis that these terms might become more popular in the fan press. We can also notice some tangible differences in the relationship between terms. *Hit* and *box office* do not share an increase in popularity as in the industry publications. The year 1937 also seems much less significant in the fan publications as *contract* had already been on the decline since 1936 and *hit* spikes at 459 page counts in 1940. It's also important to remember these results come only from a sample of the total list of terms. The lack of much terminological traffic could be a bad bet on our part, indicating that the terms we sampled did not actually move. Future studies will be able to observe the interplay using the entire list. Our preliminary effort did, however, yield one important finding.

Our tentative findings point to little variation in usage, apart from a scaling up around 1936–37 of most terms. *Gross* and *budget* are confirmed as trade-specific terms, and are not shared by fan publications. Both industrial and fan publications display considerable investment in referencing hits, and both also have considerable disinterest in referencing flops, an observation that supports the generally celebratory focus of both these publications.

A somewhat unexpected result appears with *contract*. This single term jumps out as prominent for both industrial and fan publications. This immediately suggests that here we might find a fruitful place to dig deeper, to see how the legal and labor facets of entertainment acted as a shunting point for ideas about entertainment as an industrial entity between business and general audiences. On closer study, the contracts discussed in industry publications involved all manner of relations among studios, producers, exhibitors, and talent. Those appearing in fan magazines overwhelmingly pertain to the contracts of star actors.

Coming at this from a different methodological angle, Richard Maltby published an article in *Film History* that challenged film historians to take account of contracts because they were the vehicle for the integration of local media operations with larger industrial concerns. Maltby focuses on what he claims is the most ubiquitous document of the classical Hollywood era, the Standard Exhibition Contract.[5] Our pilot study suggests that his assertions may be extended, and that contracts in general, those artifacts of business conduct, travelled to a popular understanding of the relationship between talent and the exploiters of that talent. More than that, the Arclight app usefully helped pinpoint areas deserving of more focused research attention that might have gone undetected. Taking care to explore the defining features of the corpus, we are able to identify a basis of comparison that can rattle pre-existing impressions about the operations of industry and fandom.

ENDNOTES

1 Significant thanks goes to Robert Hunt and Tyler Morgenstern for expert editorial and research assistantship in the production of this pilot study.

2 Charles R. Acland, *Screen Traffic: Movies, Multiplexes, and Global Culture,* (Durham: Duke University Press, 2003) is in part a history of the expansion of this industrial discourse into popular venues, focusing on the changes in the American exhibition business in the 1980s and 1990s.

3 See Charles R. Acland, "Senses of Success and the Rise of the Blockbuster," *Film History* 25.1/2 (2013): 11–18.

4 For more discussion, see Eric Hoyt, Kit Hughes, Derek Long, Anthony Tran, and Kevin Ponto (2014) "Scaled Entity Search: A Method for Media Historiography and Response to Critiques of Big Humanities Data Research," in *2014 IEEE International Conference on Big Data*, 51–59, http://doi.org/10.1109/BigData.2014.7004453.

5 Richard Maltby, "The Standard Exhibition Contract and the Unwritten History of the Classical Hollywood Cinema," *Film History* 25.1-2 (2013): 138–53.

BIBLIOGRAPHY

Acland, Charles R. *Screen Traffic: Movies, Multiplexes, and Global Culture.* Durham: Duke University Press, 2003.

———. "Senses of Success and the Rise of the Blockbuster." *Film History* 25, no. 1-2 (2013): 11–18.

Hoyt, Eric, Kit Hughes, Derek Long, Kevin Ponto, and Anthony Tran. "Scaled Entity Search: A Method for Media Historiography and Response to Critiques of Big Humanities Data Research." *Proceedings of IEEE Big Humanities Data. 2014.* https://bighumanities.files.wordpress.com/2014/10/hoyt.pdf.

Maltby, Richard. "The Standard Exhibition Contract and the Unwritten History of the Classical Hollywood Cinema." *Film History* 25, no. 1-2 (2013): 138–53.

DIGITAL TOOLS FOR FILM ANALYSIS: SMALL DATA

Lea Jacobs and Kaitlin Fyfe

Given the storage and processing capacities of computers, it is not surprising that one of the main thrusts of digital scholarship in the humanities is the aggregation of data. In the field of film studies, this is most obvious in the case of the Lantern search engine, which provides sophisticated tools for searching the comprehensive and growing online database of the Media History Digital Library. It is also true for the Media Ecology Project, which gives online access to moving-image research materials and seeks to facilitate connections between the archives and researchers and students. In addition, the Cinemetrics site includes software that measures shot length and other variables over the course of an entire film as well as a database of more than 15,000 titles with average shot lengths and related statistical data. But the aggregation of data is not the only way in which digital tools may benefit film and media studies. Within the film industry and the archival community, digital tools are frequently used for small-scale manipulations of the medium: editing sound and picture, color correction, restoring prints frame by frame. At the University of Wisconsin–Madison, we have been experimenting with the use of digital tools at this level, specifically with using nonlinear editing systems (NLE) to analyze the development of film style in a scholarly context (Jacobs) as well as to create teaching materials for film production students (Fyfe).

Nonlinear editing systems have several capabilities that have proven useful to our analytical efforts, including:

* close examination of the film frame by frame;
* marking the video file;
* generating visual displays of information such as charts, overlays on the image track, or some combination of the two;

* reverse engineering the post-production process by recutting scenes, altering the track, or separating out the elements of the sound mix.

The capacity to examine a film frame by frame is available to anyone working with analog film, however, we would note that you can do things with a digital file that you cannot do with a film print.

For the analysis of sound tracks, one important aspect of working digitally is the capacity to mark the sync points—to find and label the specific frame at which image and sound have been aligned. On a film print, sound and image are printed out of phase for projection. As film travels through a projector, it reaches the shutter, which displays the image, before it reaches the sound head, which relays the sound to the loudspeakers. The sound track thus has to be printed "behind" the film frames with which it is synchronized, making it difficult to ascertain the precise relationship between sound and image even when viewing a film on a flatbed editing table. In contrast, one accesses sound and image simultaneously once the digital file has been mounted in a nonlinear editing system. Moreover, the process of scrubbing (going back and forth over the sound record to locate the beginning or end point of a given sound) is not possible with film due to the damage it can cause the print, whereas video editing is nondestructive and permits this process.

BASIC WORK FLOW

Our initial experiments utilized Final Cut Pro 7. We have since made a transition to Adobe's Premiere ProCC due to the fact that Apple fundamentally altered Final Cut in ways which rendered it incompatible with our working methodology and existing project files.[1] We also make use of Excel and Photoshop. Our projects necessitate having access to a frame-accurate digital file (with all of the original film frames), which can be mounted in the editing software. The best source material is thus a progressive (frame

by frame) digital scan of a film print or a DVD which has been made by the same method and then ripped and exported into a format with minimal compression. Many older DVD transfers were produced according to the 3:2 pulldown standard for video projection in which film frames are selectively and consistently repeated to convert from cinema's 24-frames-per-second rate to video's 30-frames-per-second rate.[2] In such cases it may be possible to eliminate the interpolated frames and get back to the original using the Cinema Tools program. Sequences from all of the DVDs we work with are first examined frame by frame on a computer for the characteristic "combing" which indicates that the original film frames have been recombined with others during the film-to-video transfer. If this is the case, then we either correct the problem via Cinema Tools or relinquish the example.

Once we have ascertained that our source material is frame accurate, we mount the sections we want to analyze in the editing software. Figure 1 illustrates the workspace for Premiere; the "timeline," found on the right, is composed of the image track with related audio tracks below it.

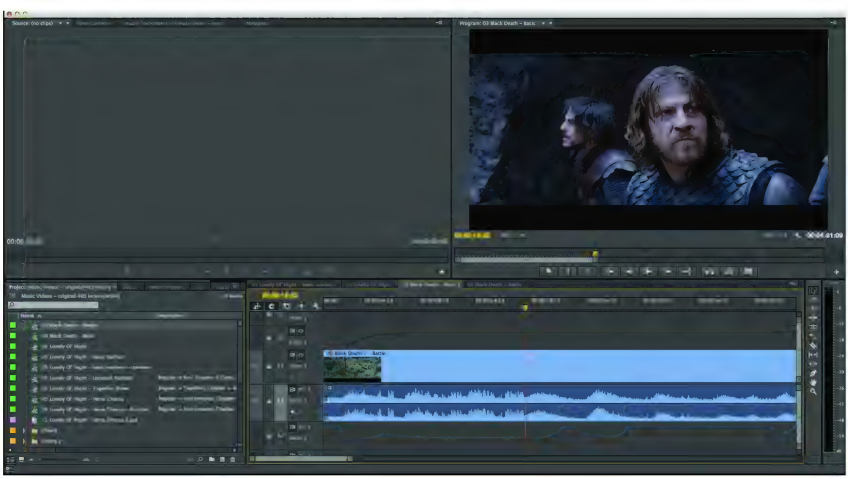


Figure 1. Premiere Pro CC workspace.

Each position, each frame within the timeline is identified by timecode, written in hours, minutes, seconds, and frames. For example, 00:03:10:05 describes a frame at 0 hours, 3 minutes, 10 seconds, and 5 frames. The software also gives us the ability to place markers throughout the timeline. The markers are like the wax pencil analog film editors used to mark a given frame of film. The digital markers we have now, however, are very flexible. They can be given a name and can store additional text in a comment field. They are navigable, providing a way to hop precisely between specific moments in a clip or sequence. They can be given a duration, marking a series of frames rather than a single frame. We can use different categories and colors of markers. They can be erased when necessary. And they can be used to extract exact timecode data about their position on the timeline.

Once mounted, a clip initially appears in the timeline as a single unbroken entity, as in figure 1. We place cuts in the clip which correspond to those in the film (see fig. 2). This facilitates navigation through the sequence, as one can move directly from cut to cut at a keystroke, play back or determine the duration of individual shots, or do the same for groups of shots.

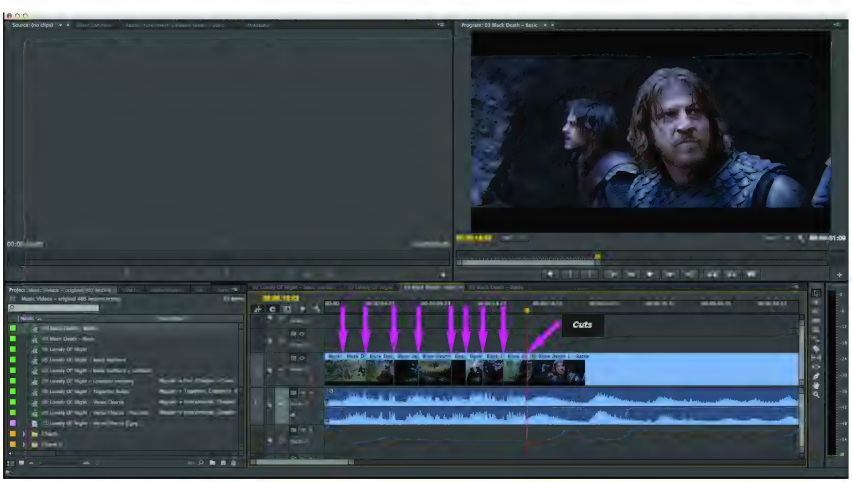
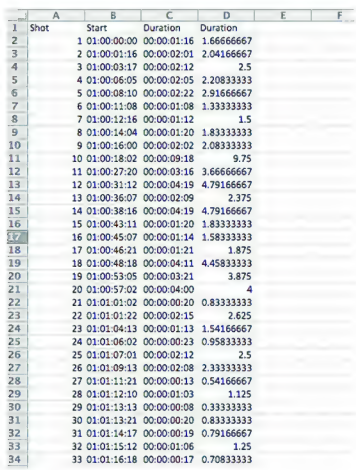


Figure 2. Master clip cut into shots.

To generate charts of shot duration, we place a marker at each cut. The timecode for all markers in the sequence can thus be exported as a Marker List, a simple CSV (comma separated values) text file that contains the timecode, name, and comments for each marker. This file can be opened in Excel (see fig. 3). We use a formula to perform some simple arithmetic on the timecode values of each marker in order to derive a duration for each shot. These values, also in timecode format, are then converted, via another formula, into seconds, with fractions of a second represented in decimals rather than frames. These values can then be used to generate a chart within Excel.



	A	B	C	D	E	F
1	Shot	Start	Duration	Duration		
2		1 01:00:00:00	00:00:01:16	1.66666667		
3		2 01:00:01:16	00:00:02:01	2.04166667		
4		3 01:00:03:17	00:00:02:12	2.5		
5		4 01:00:06:05	00:00:02:05	2.20833333		
6		5 01:00:08:10	00:00:02:22	2.91666667		
7		6 01:00:11:08	00:00:01:08	1.33333333		
8		7 01:00:12:16	00:00:01:12	1.5		
9		8 01:00:14:04	00:00:01:20	1.83333333		
10		9 01:00:16:00	00:00:02:02	2.08333333		
11		10 01:00:18:02	00:00:09:18	9.75		
12		11 01:00:27:20	00:00:03:16	3.66666667		
13		12 01:00:31:12	00:00:04:19	4.79166667		
14		13 01:00:36:07	00:00:02:09	2.375		
15		14 01:00:38:16	00:00:04:19	4.79166667		
16		15 01:00:43:11	00:00:01:20	1.83333333		
17		16 01:00:45:07	00:00:01:14	1.58333333		
18		17 01:00:46:21	00:00:01:21	1.875		
19		18 01:00:48:18	00:00:04:11	4.45833333		
20		19 01:00:53:05	00:00:03:21	3.875		
21		20 01:00:57:02	00:00:04:00	4		
22		21 01:01:01:02	00:00:00:20	0.83333333		
23		22 01:01:01:22	00:00:02:15	2.625		
24		23 01:01:04:13	00:00:01:13	1.54166667		
25		24 01:01:06:02	00:00:00:23	0.95833333		
26		25 01:01:07:01	00:00:02:12	2.5		
27		26 01:01:09:13	00:00:02:08	2.33333333		
28		27 01:01:11:21	00:00:00:13	0.54166667		
29		28 01:01:12:10	00:00:01:03	1.125		
30		29 01:01:13:13	00:00:00:08	0.33333333		
31		30 01:01:13:21	00:00:00:20	0.83333333		
32		31 01:01:14:17	00:00:00:19	0.79166667		
33		32 01:01:15:12	00:00:01:06	1.25		
34		33 01:01:16:18	00:00:00:17	0.70833333		

Figure 3. Timecode data in Excel.

In its simplest form, this gives us a shot duration chart, which we can copy and paste into Photoshop for additional annotation. Figure 4 is a shot chart for a battle scene in *Black Death* (2010), a sword and shield film.

A young monk, Osmund, leads a group of holy warriors to investigate why a town has been spared infection by the plague. In this

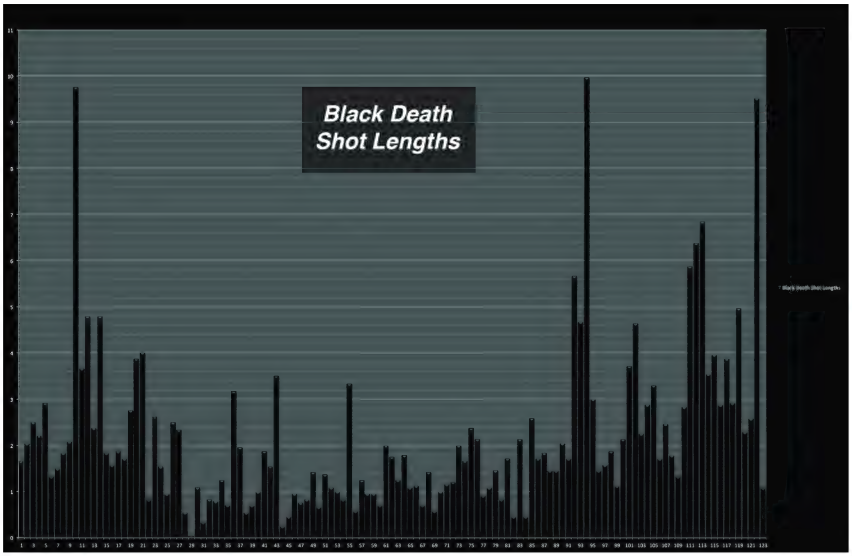


Figure 4.

sequence, they are ambushed in the middle of the forest by a large group of armed men. Figure 5 indicates the division of the scene into subsegments. The first subsegment, "Osmund Warns," is the section of the scene where Osmund runs to camp to warn everyone about the impending attack. The second subsegment, "Get Ready," is the section of the scene where the warriors ready their weapons and prepare for battle. The third and fourth subsegments focus on the battle, and so on

The segmentation was initially constructed without taking account of any formal editing patterns but rather with the aim of charting the narrative development. In this as in many other cases, however, variations in editing patterns line up with the narrative divisions of the scene. In figure 5, for example, the two battle subsegments display a faster cutting pace than the "Get Ready" subsegment. Additionally, in the second battle subsegment, "Osmund's POV," the shot length shows a tendency to increase over the course of the subsegment. This corresponds to the increasing

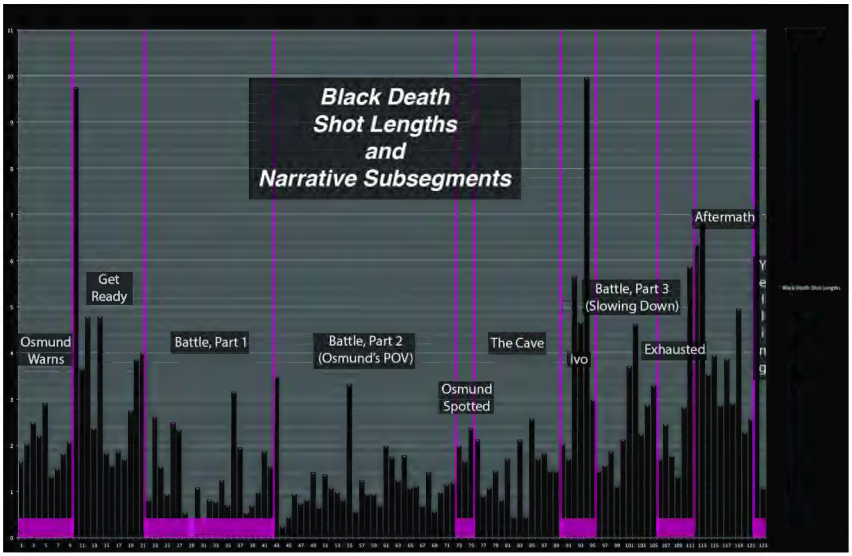


Figure 5.

focus on the brutality of Osmund's companions. A shot might be held longer to show an opponent's face screaming in pain or to show repeated hits. The chart in figure 5 also reveals a particularly slow cutting pace in the "Ivo" segment. Ivo has rescued Osmund but has been fatally wounded in the process. The longest shot in the sequence shows Ivo taking his last breath.

We have found it useful to integrate the charts of shot duration with the display of the film by pulling charts back into Premiere. This approach, shifting between chart and video in real time, helps provide an organic connection between scene and data, one that allows for a better grasp of what each moment in the chart represents and lets us explore very subtle shifts in the editing. Another advantage of putting the chart back into Premiere is that it provides us with a ready interface with which to reexamine our data and, for example, check shot lengths. In the timeline, you can see the duration of the current shot as measured by the software. Once we have prepared the clip in the timeline and the chart, we

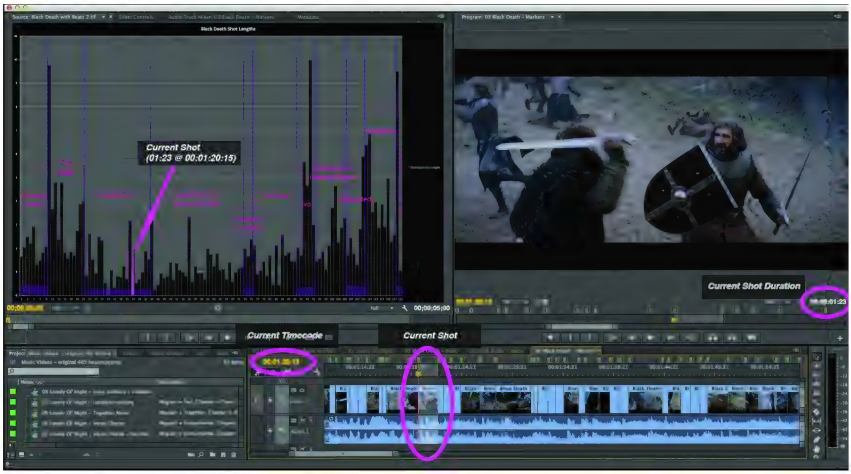


Figure 6. Shot length chart in Premiere.

can build upon already completed work in order to add additional parameters to the analysis.

Figure 7 shows a chart of shot durations for the John Mellencamp music video "Lonely Ol' Night" (1985) with additional variables, mapping out the verses, choruses, and instrumental sections (the shots containing the verses are shown in blue, the shots containing choruses in green, and the shots containing the instrumental sections in black). This chart effectively zooms in by cutting off shots that extend past nine seconds in length to provide a better view of the many shots that are below two seconds. The chart enables us to track how the cutting interacts with the verse-chorus structure of the song.

By exporting another set of markers, we can display the locations shown in each shot of the video. In the locations chart, green is used to indicate shots of a county fair, blue for shots back in town, yellow for shots of John Mellencamp performing the song on a porch, pink for shots of Mellencamp walking around with his girlfriend (either at the fair or in town), and purple for a pair of shots

of lightning in a night sky.

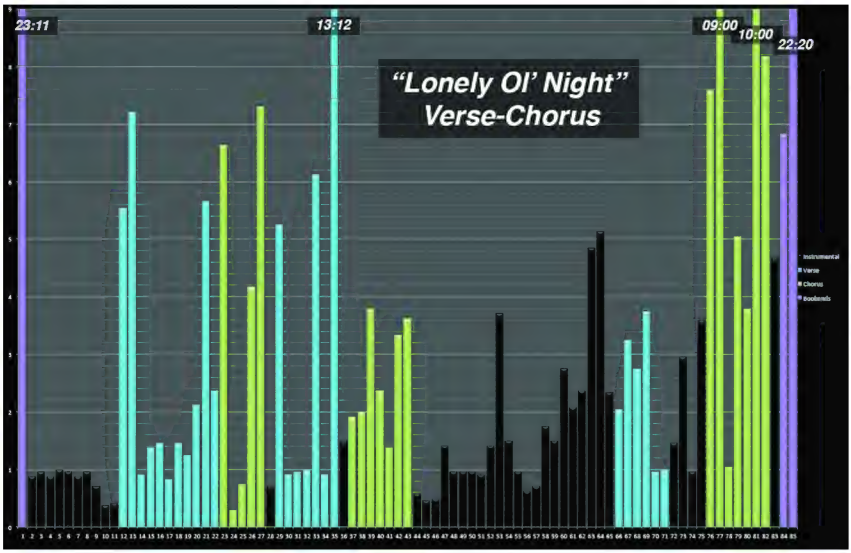


Figure 7. Shot durations with verse-chorus structure.

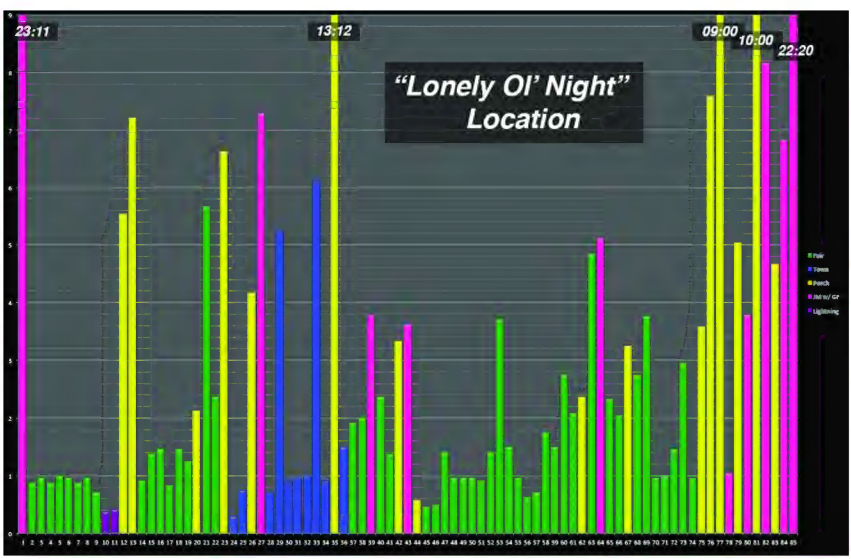


Figure 8. Shot durations with locations.

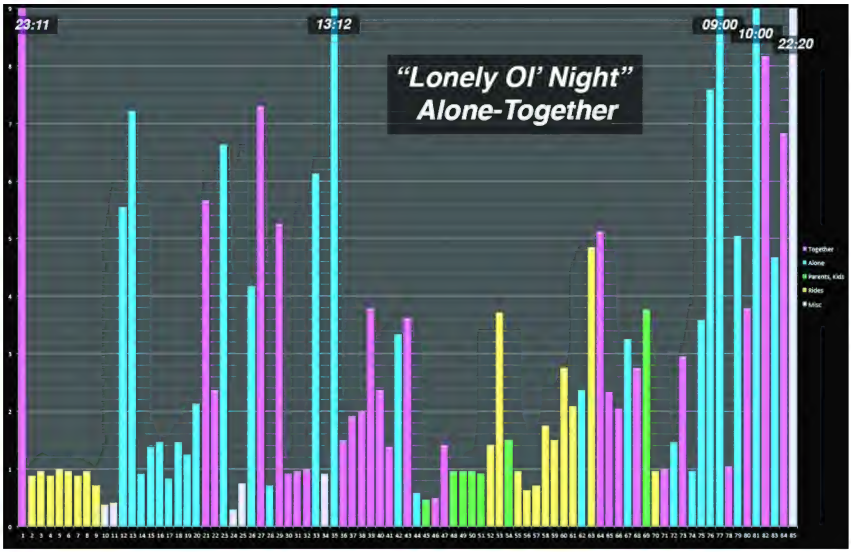


Figure 9. Shot duration with shot content.

We can also chart the subject of each shot. Figure 9 utilizes blue for shots of people who are alone, pink for shots of couples, green for shots of parents and kids, yellow for shots of rides at the fair, and light grey for shots that contain no people and no rides. We can thus relate shot duration to other aspects of the shot and explore other editing choices in addition to shot length, so that, for example, compositional elements or framing decisions can be brought into the analysis.

Finally, multiple schemas may be displayed simultaneously. Figure 10 shows location set against the verse-chorus structure.

The shot by shot categories from the verse-chorus chart have been laid out below the X-axis, lining up with the corresponding shots in the location chart. We can see how the footage of the town is introduced in the second chorus and then becomes the focus in the second verse, after which we return to the fairgrounds as our primary location. We can also see that the introduction of Mellen-

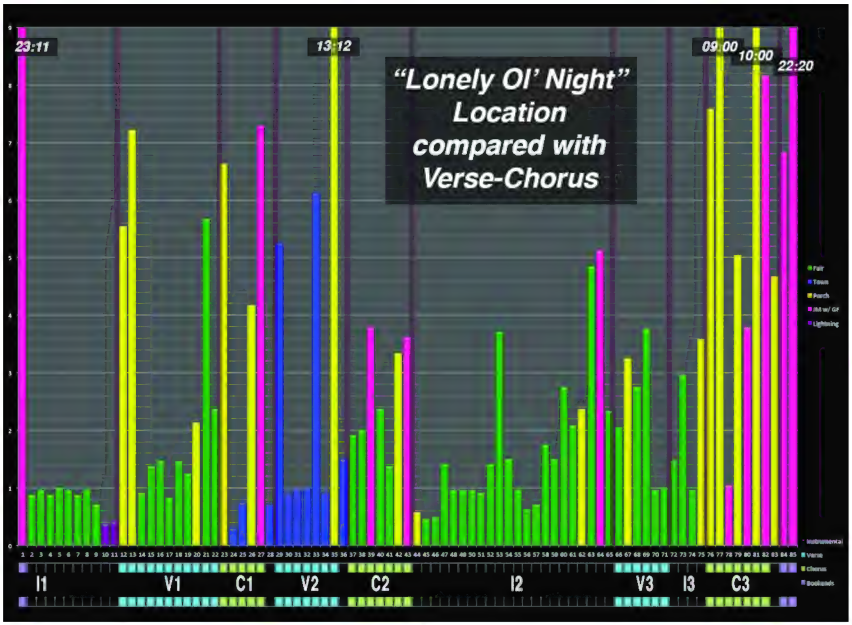


Figure 10.

camp on the porch is delayed until after the opening instrumental section, used minimally within the second instrumental section, and dominates the final chorus.

SOUND AND RHYTHM

A second set of examples derives from recent research on sound and rhythm in the early sound period.[3] One set of case studies was devoted to the early sound cartoon.

We knew that animation in the period was planned out with bar sheets. Figure 11 shows part of a bar sheet (in this cased dubbed a "lay out sheet") for *Santa's Workshop*, a Silly Symphony released by Walt Disney in 1932.[4]

Each square represents a measure, although with all musical notation removed. Note that in the case of *Santa's Workshop* the lyrics

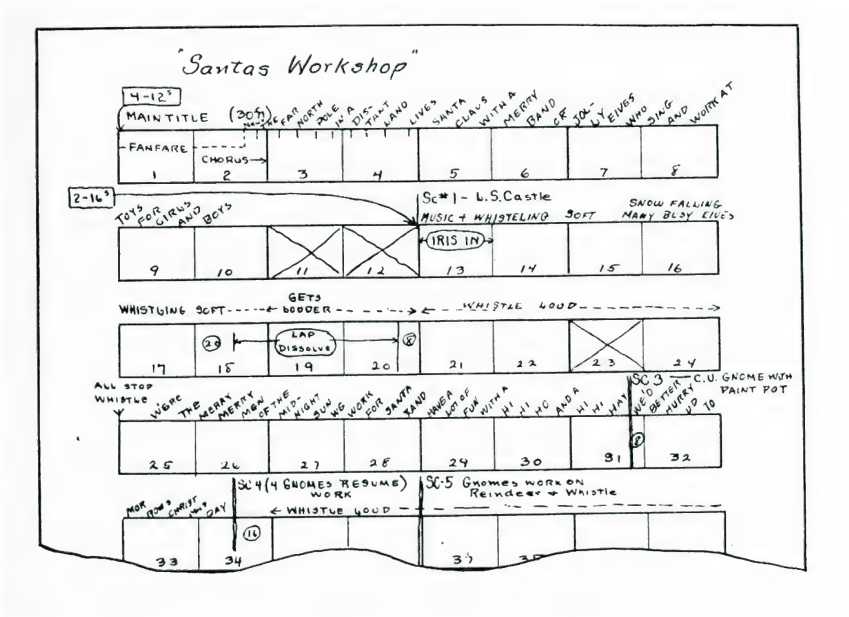


Figure 11. Bar sheet for *Santa's Workshop*.

are written above the bar, indicating the placement of words in relation to the beat (dialogue was handled the same way). The length of the lap dissolve in bars 18–20 is indicated in frames. Similarly the location of cuts is precisely specified. For example, the cut in bar 31 comes eight frames before the end of the measure. More generally, the timing of the animation and tempo of the action is indicated by figures at the beginning of each section, showing the number of beats per bar and number of frames per beat.

We set out to reconstruct the digital equivalent of these bar sheets for a number of Disney cartoons. Take, for example, a gag from *Playful Pluto* (1934) in which a wind vortex, having taken up a pile of Mickey's carefully raked leaves, deposits them in a basket only to have Pluto rifle through and scatter them in search of a bone. The analysis (done in Final Cut) was initiated by marking each

beat and bar for the segment, as in figures 12–15.

Figure 12. Downbeat of bar 60.

Figure 13. Downbeat of bar 61.

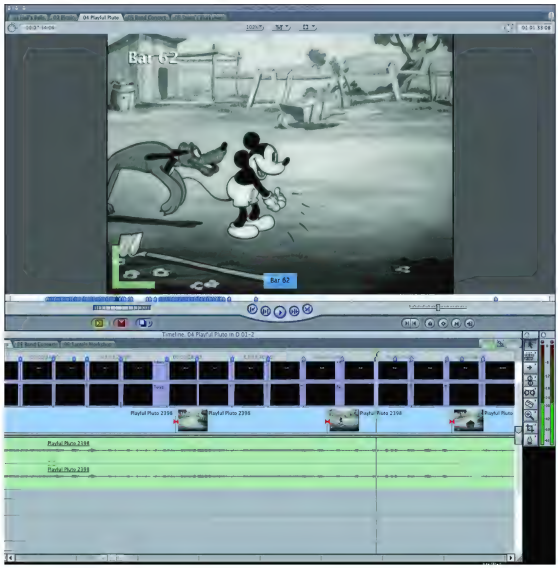


Figure 14. Downbeat of bar 62.

Figure 15. Third beat of bar 62.

This system of marking the sync points made it possible to notate cuts and shot descriptions directly on the score.

Note, for example, that the cut to Mickey in shot 9 occurs in bar 60, while we still have the triplets associated with the wind vortex. Much of the subsequent action is timed to coincide with the strong first and third beats of the 4/4 meter. The animators prepare for

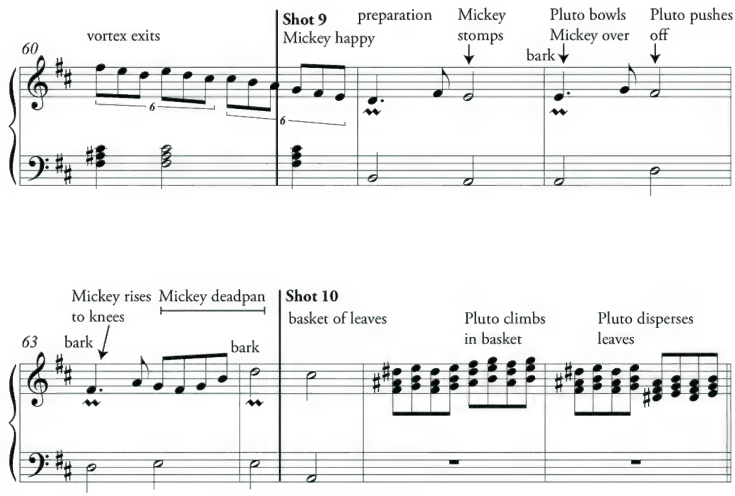


Figure 16.

Mickey's stomp over the downbeat of bar 61 and the stomp actually occurs on the third beat. Pluto enters on the downbeat of bar 62 and exits on the third beat. Note the length of time that the animators have allowed for Mickey's deadpan (16 frames) as he looks off frame watching Pluto prepare to scatter the leaves that the vortex has deposited in his basket.

We were able to pull this chart back into Final Cut Pro (FCP) to generate a clip that included picture, sound, and written score as demonstrated in clip 1. We utilized the FCP beat markers to produce and export a MOV file with frame-accurate indications of

the bar changes appearing in the overlay on the upper left corner of the frame. We also animated a pink sliding bar on the score to the same markers. This system permitted a visual representation of how the timing of the animation was integrated with the music track. In addition, we discovered further rhythmic dimensions of sound effects such as Pluto's barks on the downbeat of bars 61–63. The research on sound and rhythm extended to live action films, including dialogue timing and performance. Complaints about slow and draggy dialogue scenes were frequent in the popular and trade press during the early sound period.[5] Case studies examined a number of films made by Howard Hawks between 1930 and 1934. Editing software allowed us to mark the starting point of words or word fragments and to measure the duration of utterances or silences to the frame. We could thus analyze dialogue according to the following parameters:

* tempo of line readings (measured in words/second or words/frame);
* pauses between words or phrases;
* dialogue overlaps between speakers;
* how much of a given word or phrase overlapped a cut (shot overlaps).

It was also possible to gauge with precision the relationship between an actor's words and the starting or end point of movement and gestures. Thus, we could pin down the temporal relationship between speech and gesture, speech and cuts, and, in the case of dialogue underscoring, individual words and notes in the underscore. Take this interchange from the 1932 film *Scarface*, clip 2.

The actors each have distinct styles of line delivery, and over the course of the clip the advantage switches from Osgood Perkins, playing Lovo, to Paul Muni, playing Tony. Lovo is voluble and Tony is laconic. Perkins as Lovo speaks quickly, at rates over five words per second. His relatively long lines, peppered with color-

ful metaphors, are interrupted by his interlocutor's brief interjections. His dialogue builds in speed and intensity over the course of shot 4, the longest in the sequence, at 27:18. In addition to the sheer volume and speed of the words, dialogue overlaps and camera movement also favor Perkins's speech. As Muni finishes his first short line, Perkins steps on it very briefly, rushing in with the word "Now" (highlighted in yellow). He stands up speaking as the camera moves in (highlighted in blue): "Now you listen to me stupid. That was one of O'Hara's places and you know it." Even after he has completed the action of standing, the track-in continues, accenting the phrase: "Didn't I tell you I wasn't ready for O'Hara yet?" As the conversation continues, Perkins again steps on Muni's line (highlighted in yellow), "Ah, don't worry," with the word "Don't" leading to his longest speech, delivered with lightning speed and a distinct internal rhythmic structure based on the repetition of questions ("Don't worry? You know what O'Hara's liable to do now?"), the repetition of phrases ("You're liable to get it, and I'm liable to get it"), and the comparison, which the stressed syllables emphasize, between "guns" and "hummingbirds."

Shot 4. Medium long shot with Tony standing left and Lovo seated right.

TONY [3.6 words/sec.]: Listen, Johnnie, it was easy. [Dialogue overlap, one frame, "–y" in "easy" and Lovo's "N–" in "Now"]

LOVO [5.1 words/sec.]: Now you listen to me, stupid. [Perkins starts to stand up and camera starts to track in.] That was one of O'Hara's places and you know it. [He is standing.] Didn't I tell you I wasn't ready for O'Hara yet? [Track-in complete.]

TONY [4.0 words/sec.]: It was a nice little order. Fifty barrels a week.

LOVO [5.4 words/sec.]: I don't care if it's fifty a day. What do you use to think with, an empty beer keg? Just when we get this territory lined up and

runnin' smooth, you step out and gum up the parade.

TONY [3.9 words/sec.]: How do you mean, Johnnie? I just sell a little more beer. Ah, don't worry. [Dialogue overlap, one frame, –y in "worry" and D– in "Don't"]

LOVO [5.3 words/sec.]: Don't worry? You know what O'Hara's liable to do now? He'll send his guns down here on the South Side, they'll move around like hummingbirds. You're liable to get it, and I'm liable to get it. I know that hop, he's tough, see [shot overlap, one frame past cut].

Muni says less and speaks relatively slowly, with a spread of speaking tempi from 1.3 to almost 4.0 words per second. Muni's performance also makes significant use of pauses, one of which is underlined by the editing. The line "Ahhhh, he ain't so tough, you afraid of a guy like that?" is edited to put the word "that" after the cut to shot 6 (highlighted in blue), so that we can see Perkins's quick reaction to the implication that he is a coward. The whole pace of the scene then slows for a double pause. Following Perkins's denial, "I ain't afraid of anybody," there are almost two seconds of silence that overlap the cut back to Muni's close up as the actor waits to respond. His ironic response is divided by another break between the slurred phrase "Surenot" and "That's a crazy question, eh, Johnnie?"

Shot 5. Medium close up, Tony, full face.

TONY [2.4 words/sec.]: Ahhhh, he ain't so tough. Hanging out in a flower shop. You afraid of a guy like

Shot 6. Medium long shot, back of Tony, favoring Lovo.

TONY: that? [Cut divides the words: "like" / cut / "that?"]

```
LOVO: I ain't afraid of anybody. [Pause for 01:20.]
    Shot 7. Medium close up, Tony, as 5.
TONY [1.3 words/sec.]: Sure not. [Words slurred. Sec-
    ond pause.] That's a crazy question, eh, Johnnie?
```

It may seem improbable that actors, directors, and editors could or would seek to control the timing of speech and gesture to this degree, but many accounts by practitioners suggest this is the case. For example, in his treatise on film editing Karel Reisz notes:

> [T]he editor must respect the actor's performance. In an action scene, the exact timing of shots is very often left open to the editor and he can impose a pace on the sequence which he considers most fitting. In a passage of dialogue his problem is more complex because an actor sets his pace in the playing. If the editor wishes to speed up the continuity, he can shorten the pauses between sentences, use cheat cuts and generally cut down all the footage not "anchored" by the dialogue. . . . But interfering in an actor's performance can sometimes cause more harm than good. An experienced actor with a developed sense of timing may set his own pace during a scene which it is best to leave alone. . . . The moments preceding and following the actor's words are an integral part of his interpretation of the line, and to eliminate them may reduce the effect of the rendering.[6]

The challenge for the editor is to create smooth and visually compelling matches that respect the timing of the actor's performance. The editor's decisions, like the actor's, will involve very small durations: cutting on discrete phonemes or on the time of the actor's breath, and placing the picture edit precisely in relation to changes in gesture and facial expression.

Our work points to the importance of considering film construc-

tion at a very small scale. In the case of the analysis of editing patterns in *Black Death* and "Lonely Ol' Night," we sought to understand changes in shot duration over relatively brief segments as opposed to the analyses made using the Cinemetrics tool cited above, which are typically concerned with average shot lengths over an entire film or group of films. In addition, the investigation of the role played by music in timing animation and of dialogue timing and performance calls attention to the close management of sync points at the level of the beat or the word. While one might think that such fine-grained analysis would only be fruitful in very restricted instances, such as animation, our work suggests that live-action sound and picture editors also regularly deal with very small durations, planning their work from sync point to sync point and frame to frame. Digital tools have permitted us to understand the management and control of timing at this level. Thus, in addition to the "big data" commonly associated with digital humanities, we would assert the interest of "small data." Like the microscopic examination of a tissue sample, the close examination of film via digital editing systems opens up a whole new research perspective.

ENDNOTES

1 For more discussion of the changes Apple made to Final Cut and its impact on the post-production community, see Mark Raudonis, "'Real World' Editing: From Avid to FCP and Back Again," CreativeCOW.net, accessed March 1, 2016, https://library.creativecow.net/raudonis_mark/magazine_29_Real-World-Editing/1 and David H. Lawrence, "The Magnetic Timeline: Thoughts on Apple's New Paradigm, Revisited," CreativeCOW.net, accessed March 1, 2016, https://library.creativecow.net/lawrence_david/Magnet-Timeline/1 .

2 See David Bordwell, "My name is David and I'm a frame-counter," *Observations on Film Art,* accessed April 6, 2016, http://www.davidbordwell.net/blog/2007/01/28/my-name-is-david-and-im-a-frame-counter/.

3 Lea Jacobs, *Film Rhythm After Sound: Technology, Music and Performance* (Berkeley: University of California Press, 2015).

4 William Garity, "The Production of Animated Cartoons," *Journal of the Society of Motion Picture Engineers* 20, no. 4 (1933): 309–22. Garity's bar sheet is reproduced here as figure 14.

5 For a sense of this discourse, see Jacobs, *Film Rhythm*, 3–9.

6 Karel Reisz and Gavin Millar, *The Technique of Film Editing* (first printed 1953; reprint London: Focal Press, 1968), 100.

BIBLIOGRAPHY

Bordwell, David, "My name is David and I'm a frame-counter." *Observations on Film Art.* Accessed April 6, 2016. http://www.davidbordwell.net/blog/2007/01/28/my-name-is-david-and-im-a-frame-counter/.

Garity, William. "The Production of Animated Cartoons." *Journal of the Society of Motion Picture Engineers* 20, no. 4 (1933): 309–22.

Jacobs, Lea. *Film Rhythm After Sound: Technology, Music and Performance.* Berkeley: University of California Press, 2015.

Lawrence, David H. *"The Magnetic Timeline: Thoughts on Apple's New Paradigm, Revisited."* CreativeCOW.net. Accessed March 1, 2016. https://library.creativecow.net/lawrence_david/Magnet-Timeline/1.

Raudonis, Mark. "'Real World' Editing: From Avid to FCP and Back Again." CreativeCOW.net. Accessed March 1, 2016. https://library.creativecow.net/raudonis_mark/magazine_29_Real-World-Editing/1.

Reisz, Karel and Gavin Millar. *The Technique of Film Editing.* London: Focal Press, 1953. Reprint, London: Focal Press, 1968.

THE SLICES OF CINEMA: DIGITAL SURREALISM AS RESEARCH STRATEGY

Kevin L. Ferguson

No doubt the most famous slice in cinema is to be found in Luis Buñuel's short film *Un Chien Andalou* (1929), which opens with a prologue where a man slices a woman's eyeball with a straight razor.[1] Shocking, violent, disgusting: Buñuel's film remains an important landmark in cinema history, and yet the larger aesthetic and critical tradition it belongs to—surrealism—was soon mostly ignored in favor of the conservative, dominant, narrative form of cinema we know today. I lament this; like the surrealists, I search for something in the cinema that I cannot otherwise see, something that is present but hidden by my rational mind. I want to see something new, having already seen what is old.

In this, I am inspired by the avant-garde film criticism practiced by scholars such as Robert B. Ray and Tom Conley, who not only express a skepticism toward dominant film theory's rationalist, positivist, semiotic bent, but also encourage an embrace of speculative methods that stem from an often dismissed Impressionist-Surrealist tradition. Thus, Conley introduces his *Film Hieroglyphs* with an appeal "to the delineations and confusion of images and writing in the tradition of surrealist painting and cinema" and Ray explains *How a Film Theory Got Lost* by showing how "the crossroads of magic and positivism" that defined early cinema became a "spell" that traditional film criticism would later try to break.[2] So, to take one example, the faith in chance offered by the now abandoned theory of *photogénie* in the 1920s was defeated in the 1950s by the rigor of control espoused by the now canonical *politique des auteurs*. In this way, the positivist, empirical trajectory of the majority of film criticism has limited the options for media historical analysis and research.

The return to a surrealist and avant-garde tradition requires a unique kind of research, which is newly possible now that media historians have made the digital turn. While skepticism over the positivist nature of computer-based scholarship has led to critique of the digital humanities as leaning too heavily on "research instruments [that] have been absorbed from disciplines whose epistemological foundations and fundamental values are at odds with, or even hostile to, the humanities," I argue that there is a space for surrealism as a method even within empirical tools, such as the scientific image-analysis software I describe below.[3] Historically, media scholars have turned to the archive for the "hidden" contexts of media texts, such as uncovering distribution networks, narratives of collaboration, or production budgets and funding sources. Yet, rather than only take "hidden" in the sense of being buried or difficult for the researcher to find, I take a surrealist view of the hidden in order to imagine what aspects of media texts are literally impossible to see without special computer-assisted techniques. What in the archive is in plain sight but still invisible? What in the cinema is so buried that our naked eyes are unable to see it?

My approach to this question begins from Buñuel's slice, cutting the moving image's temporal tyranny and reconstituting those slices in transformative new ways. Such an approach to the visual image offers a venue for critical and aesthetic meaning-making, but it also opens new avenues for historiographic research. In the following, I define my strategy of "digital surrealism," briefly demonstrate techniques enabled by slice-based manipulations, and suggest some preliminary problems for media historical research which these techniques can address. In doing do, I hope readers can see how recalling a lost avant-garde tradition might help us productively reimagine the boundaries of scholarly media-historical analysis and research.

DIGITAL SURREALISM

Mary Ann Caws begins her comprehensive survey of surrealism by invoking its practitioners' central desire: "to turn the alert, thinking being over to the illogicalities of chance . . . [to] free the self from its logical restrictions."[4] Roland Barthes, in his early sixties structuralist phase, also recognized how the stumbling block of logic had restricted critical endeavor. Inspired by the surrealist refusal to distinguish the plastic arts from poetic language, Barthes imagined that critical investigation of the world was the inverse of artistic creation: both artist and critic represent the world by a "directed, *interested* simulacrum" of nature that "makes something appear which remained invisible or . . . unintelligible in the natural object."[5] This "structuralist activity" shared alike by artists and critics draws out the otherwise hidden present, making legible that which was illegible before. Furthermore, the activity of creating an interested simulacrum is not just distanced, objective theorizing, but rather real labor. Such creation is a meaning-making activity that ultimately betrays the critic and her world: "the simulacrum is intellect added to object, and this addition has an anthropological value, in that it is man himself, his history, his situation, his freedom, and the very resistance which nature offers to his mind."[6] What else do we see in the web browser but the anthropology of our history, our situation, our freedom, our resistance to our computer?

Digital media practitioners today experience the same, and the traces of their labor and tools are even more visible, often intentionally so given the digital humanities' spirited impulse to collaboration. The "anthropological value" of digital media practices today is very high, particularly past the digital turn where we deal not only with the "resistance which nature offers" to us, but also the resistance of the digital tools and institutions with which we wrangle the natural and its media historical expressions. Barthes's five-decade-old descriptions of structuralist activity surprisingly surface today in the work of celebrated digital humanists like

Franco Moretti, whose *Graphs, Maps, Trees* shows readers how to transform a text into "a new, *artificial* object" in a process that is nearly identical to Barthes's "directed, *interested* simulacrum."[7] Indeed, Moretti's book can be read as an articulation of the different forms of "interested simulacra" that Barthes first proposed in 1963.

Barthes should not be considered a surrealist, but numerous scholars have pointed out how his account of the "third meaning" of cinema is indebted to surrealism. Ray, for example, sees Barthes as simply "converting [surrealist] *fetishism*, with its overvaluation of apparently trivial details, into a research strategy."[8] Film scholar Adam Lowenstein has more fully drawn the history of Barthes's relationship to the surrealists, but of particular note to my project is the moment when Barthes breaks from them. Lowenstein shows how both Barthes and the surrealists "tend to focus on a certain detail in the cinematic or photographic object that 'pricks' them, unleashing a deeply felt but idiosyncratic 'spark' or 'floating flash.'"[9] Yet Barthes, who called this effect the *punctum*, was never able to accept that it was possible in the cinema: "in front of the screen, I am not free to shut my eyes . . . I am constrained to a continuous voracity; a host of other qualities, but not pensiveness."[10] But what if Barthes were not constrained to cinema's "continuous voracity"; what if he were allowed pensiveness in front of the film? The digital surrealism I propose may not address all of Barthes's prickly concerns, but it does allow critics new opportunities to generate surrealist "floating flashes" and "pricks" from the moving image.

Barthes's well-known indifference to the cinema came from a frustration with analysis; like the surrealists, he found popular cinema to be too oppressively logical and narrative for a spectator to gain any genuine shock of pleasure. Surrealist viewing practices reflect these concerns by invoking chance, play, and interaction in the search for "irrational knowledge," as with André Breton's

practice of jumping between film screenings as soon as he got bored or the "irrational enlargement" game that isolates an object and poses questions of it. But one surrealist practice in particular speaks to a digitally informed media studies: Salvador Dalí's paranoiac-critical method, an attempt at irrational knowledge that springs from unexpected juxtapositions of unrelated elements. Scholar Paul Hammond explains how "paranoia-criticism thrives on contrived delusion, on the assiduous ambition to get things wrong, to see something as something other."[11] To see something as something other; I take this as my charge. If Barthes can be read as a proto–digital humanist, we might today put him back on a surrealist trajectory that would arrest film's temporality and separate narrative logic from other forms of knowledge.

And a digital surrealism? I imagine here a range of games played on a computer with the raw material of media. Caws notes how "for years, anyone involved in Surrealism placed . . . an almost unlimited faith in automatic processes of many kinds."[12] Many of these processes involved manipulating the mechanical or technical methods of creating art, whether poetry, drawing, photography, or cinema. For example, the most striking example from her book is a reproduction of Raoul Ubac's "Fossile de la Tour Eiffel" (1938–39), a bas-relief photograph created in the darkroom by printing sandwiched positive and negative prints slightly out of alignment, which today can be reproduced in Photoshop using simple, automatic processes like high-pass and emboss filter menu actions (figure 1: Ubac on the left, a Photoshop approximation on the right). Exploring the strange possibilities of modern software—what does clicking here do?—accelerates the method-based experimentation of surrealists like Ubac and creates new opportunities for irrational knowledge that sees something as something other. Thus, digital surrealism follows in the formal tradition of surrealism by favoring automatic methods and pursuing these in a controlled, systematic way with the purpose to uncover knowledge not immediately perceptible to the rational mind.

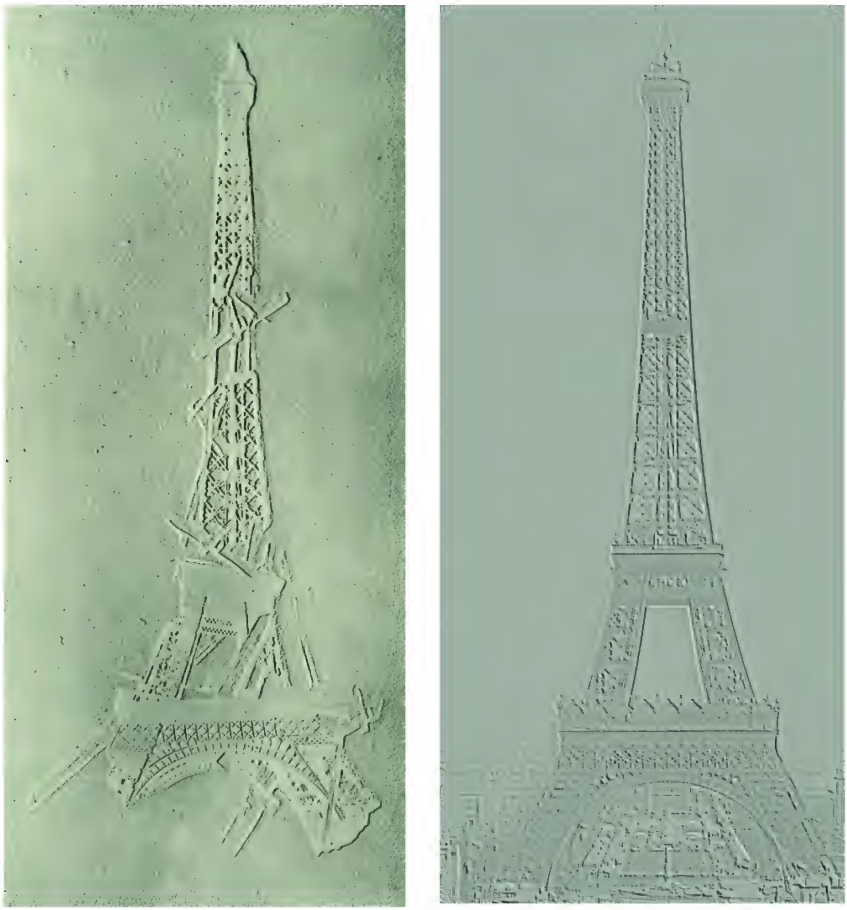


Figure 1. Ubac, "Fossile de la Tour Eiffel" and a Photoshop version.

THE SLICE

My work is based on the slice, the shuttering, juddering treasure snatched from the temporal flow of the film. Inspired by Buñuel's film, I propose "slicing" our vision, a method of digital surrealism that transforms the media text into something wholly new as an object of investigation by first cutting it to pieces. Creating a digital slice is a different conceptual process than cutting the analog filmstrip into frames. For my work, I make use of Quicktime Player 7's Export feature, which provides a "movie to image se-

quence" option to create a series of images at an interval specified in frames per second (this option was removed in later versions of Quicktime). These can be fractional, so a setting of "0.10 frames per second" results in a folder of one slice every six seconds; for a ninety-minute film, this would yield an evenly distributed sampling of 900 frames (fig. 2: *Casablanca* [1942] in 922 evenly distributed slices).

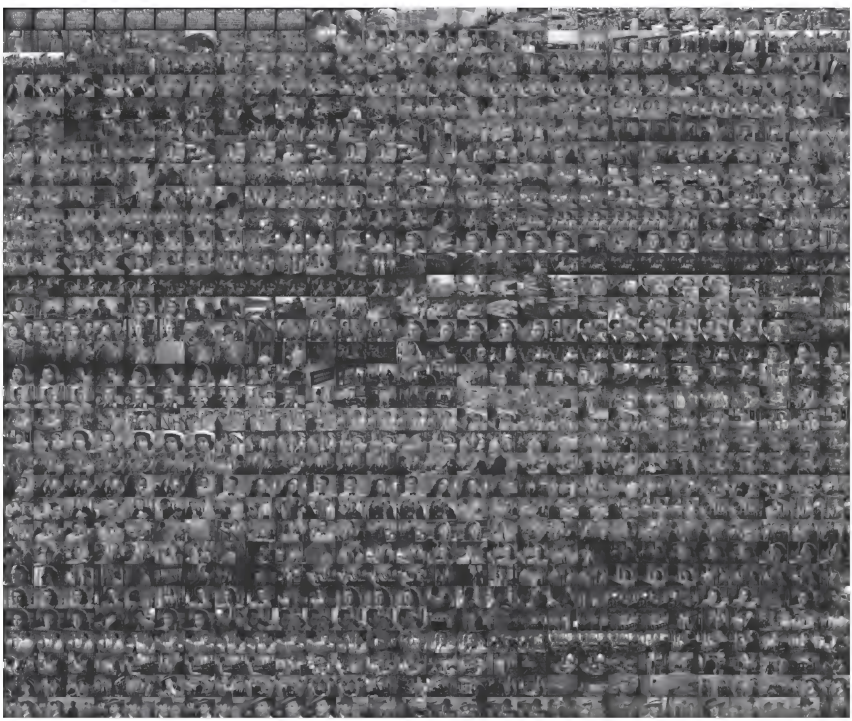


Figure 2. 922 evenly distributed frames from *Casablanca*.

The slice is indiscriminate: it does not consider narrative logic, aesthetic beauty, or the conventions of the pause. Occasionally black slices will line up with scene transitions, but most often slices appear as mid-pan blurs or with faces stuck in awkward grimaces. The slice gives unequivocally equal weight to every moment of the film, allowing the possibility that every background

object can occupy a central role. Thus, the first step in seeing the film as something other is to scroll through the slices at random, peeking in from a distance at strange shapes or colors and irrationally enlarging the found objects. We can also use the slice to move transversely through a corpus of films. What happens in the thirteenth minute of each of the 54 feature films produced by Walt Disney Animation Studios? We can easily cut across our folders using search (13*60*fps) to extract the appropriately numbered slices (fig. 3).

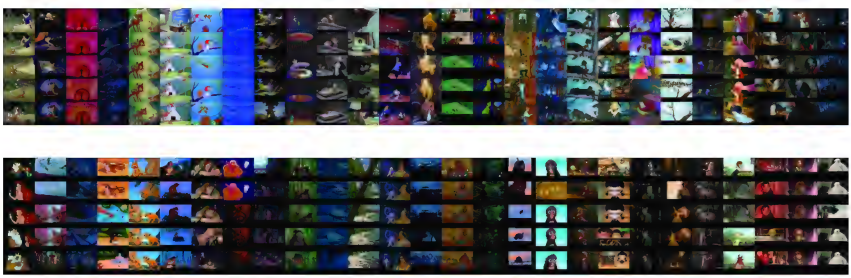


Figure 3. Frames from the thirteenth minute of 54 Disney films.

To manipulate collected film slices I use ImageJ, a public-domain scientific image-analysis software that can perform a variety of image processing tasks on a range of media formats. ImageJ's open-source framework also allows users to write and freely distribute their own Java-based plugins and macros, which Lev Manovich's Software Studies Initiative in particular has used to create visualizations of large corpora of manga, magazine covers, and Instagram photos. While ImageJ can handle video formats like AVI, doing so requires a great amount of processing power for large files such as feature-length films. The evenly distributed slice solves this problem. In scientific and medical fields, ImageJ is often used to visualize otherwise impossible-to-see structures, such as the interior of the human body or the structure of microscopic cells. Scientists first capture a series of two-dimensional images of the desired object and then "stack" them up to create a three-dimensional representation. For microscopic structures this

means taking photographs at varying focal depths; for internal structures it requires cross-section scanning like an X-ray. In both cases it is preferable for the scanned object to remain still. Applying this technique to a motion picture gives us the opportunity to also visualize the dimension of time. Elsewhere I have described a volumetric analysis of cinema where, using stacks of slices in ImageJ, I treat cinema and media texts as something like temporal tomograms, slice-based volumes that transform the dimension of time into a third, spatial dimension.[13] This third z-dimension unravels the filmstrip to transform it into a cube (fig. 4: *Casablanca* as a cube). Thus, we can spatialize time, turning the film or film scene around as an object to better see its structure, internal shape, and patterns. Most importantly, we can begin to break some of our own perceptual habits and cultural assumptions about the depth of moving images. But because we are working with a three-dimensional volume, these techniques are best suited to hands-on experimentation; unlike a bivariate graph, three-dimensional film volumes require play, manipulation, experimentation.

Figure 4. *Casablanca* visualized as a cube in ImageJ.

Barthes again: "Structural man takes the real, decomposes it, then recomposes it; this appears to be little enough . . . [y]et from another point of view, this 'little enough' is decisive: for between the two objects, or the two tenses, of structuralist activity, there occurs something new."[14] In this vein, the work I describe is creation as much as discovery, and this creation requires unexpected transformations from the algorithmic unconscious. But rather than propose a return to modernist fantasies of the recovery of a true, deep interiority—a privileged unconscious waiting to be uncovered with just the right amount of archaeological excavation—I desire instead to access a machinic unconscious. When I (improperly) use scientific image-analysis software to look at narrative moving images from radical perspectives, I gain access to a substitute, surreal, algorithmic unconscious which can be read in new and productive ways. If surrealism is, in part, about alienating the rational mind in order to produce new insight, then digital surrealism is about having machines think alongside us so that we might adopt their surreal perspectives as our own.

While not always drawn together, there is a tradition of experimental work such as mine that balances between both new media art and digital humanities scholarship. For example, since the mid-1990s new media artists with ART+COM Studios have been working with slice-based material to create "parametric translations of movies into space." To draw such experimental, media-based manipulations into more concrete theorizing of the digital humanities (DH), we could observe how Lisa Samuels and Jerome McGann's 1999 essay "Deformance and Interpretation" has influenced a large number of contemporary DH scholars, such as Stephen Ramsay, Julia Flanders, Bethany Nowviskie, and Mark Sample. Samuels and McGann suggest "deformance" as an interpretive strategy that shows how "'meaning' in imaginative work is a secondary phenomenon, a kind of meta-data [that] is important not as explanation but as residue."[15] Inviting us to

rethink "conceptual," meaning-seeking interpretations of art by considering the alternative kinds of knowledge created by "performative" operations, Samuels and McGann's deformance of poetry calls on a tradition of humanistic inquiry that makes more visible the interpreter's role and point of view in creating meaning. More than playful remix or mashup, deformative readings of aesthetic works produce serious interpretive lines of thought out of an explicitly experimental, performative framework, rather than a pre-decided, structured, theoretical concept. Thus, for the media historian expecting an explicitly structured tool to distant read corpora of film texts, the experiment of digital surrealism I propose may appear insufficient. But seen in the context of new media art and deformative DH, the researcher's idiosyncratic performance of digital surrealism can result in "the dramatic exposure of subjectivity as a live and highly informative option of interpretative commentary" otherwise neglected by "neoclassical models of criticism that search imaginative works for their 'objective' and general qualities."[16] As a popular example of the latter, consider the Cinemetrics approach, which proceeds from a very particular interpretive framework in order to analyze films by counting and comparing shot lengths. (Samuels and McGann would no doubt point out how the performance of timing shots of a film is itself rhetorical in unacknowledged ways.) My aim here is not to replace those models of criticism, but to augment them and to illuminate their shortcomings. Following Samuels and McGann, I hope media historians can see how digital surrealism, like deformance, aims to embrace and explore methodological subjectivity, rather than to pretend it does not exist.

SUMS

Having shown earlier how the film slice can be extended into a third spatial dimension, I now want to pursue in more detail a second surrealist-inspired approach to a digital humanities–influenced media studies, which follows the analogy of "distant reading" in literary studies: is it possible to view a set of feature-length

Figure 5. Summed and individual frames from *My Dinner with Andre*.

films from a distance, looking at the corpus indistinctly? Using ImageJ's z-projection feature, I create "sums" of films by adding a stack of individual slices to each other in order to create a new image. While a few of these summed films have recognizably human shapes, such as the sum of *My Dinner with Andre* (dir. Louis Malle,

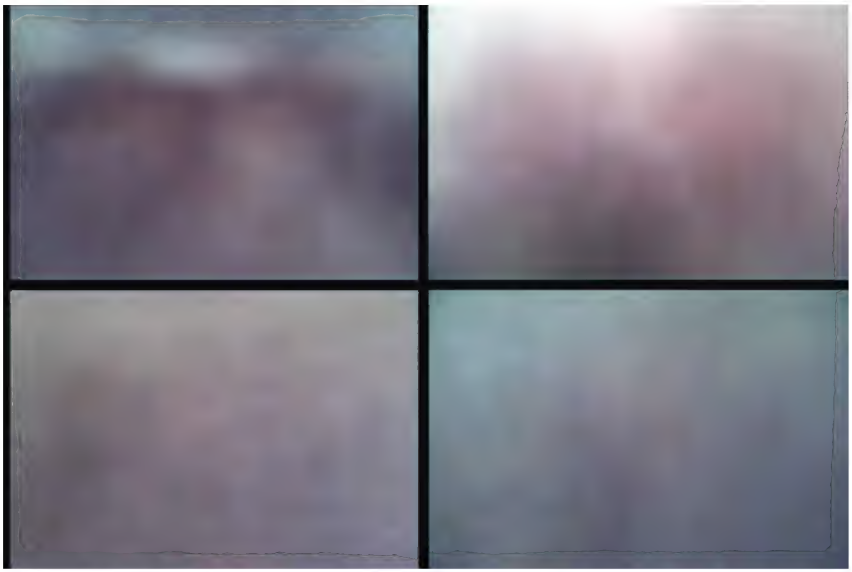


Figure 6. Summed frames of *Timecode*.

1981), where the repeated close-up on Andre's face is clearly discernible (fig. 5), the majority of them look like the sum of Mike Figgis's complexly interwoven experimental split-screen *Timecode* (2000), which shows the more common, fluid consistency of its world with only subtle variations in the bluer bottom right frame (fig. 6).

Normalizing summed frames of a larger corpora of films to a 1.33:1 "Academy" aspect ratio allows for a media historian to gain a quick visual comparison of color, intensity, brightness, and shape. As an initial point of comparison, I created four corpora of what would normally be considered rather different types of films: (1) the animated features produced by Walt Disney Animation Studios, (2) a representative selection of the western genre (including American and Italian "spaghetti" westerns), (3) a group of gialli (stylish horror films originating from Italy that influenced American slasher films), and (4) the series of popular Japanese *Zatoichi* films, following the adventures of the titular blind mas-

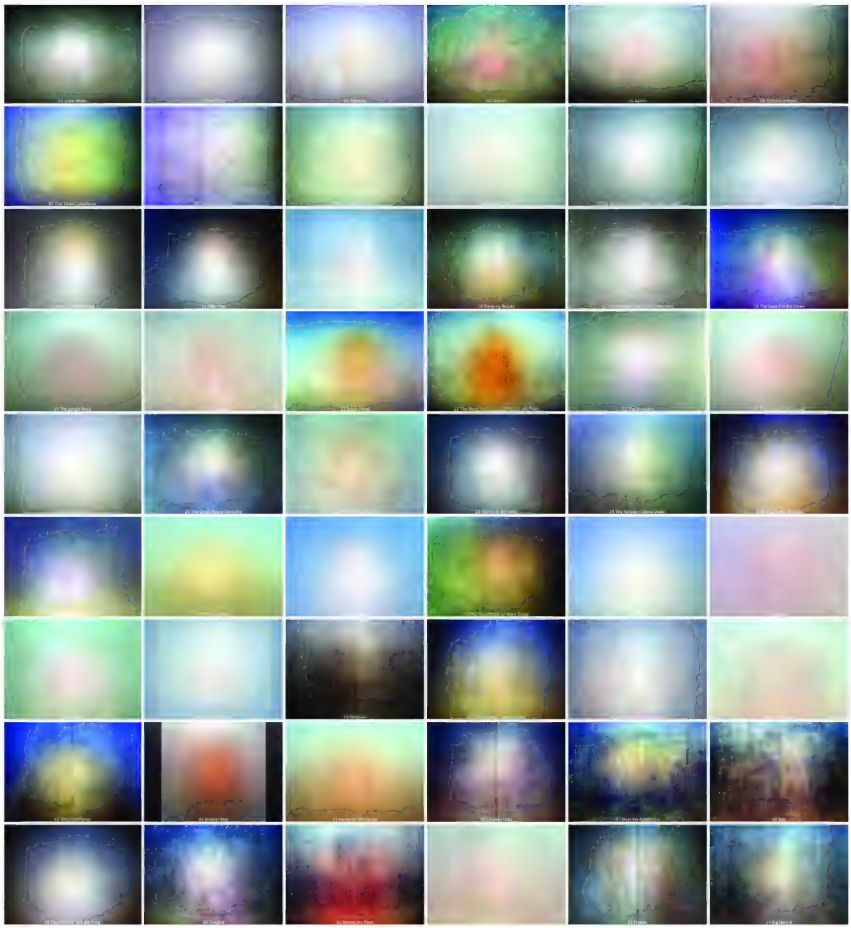


Figure 7. Summed frames of 54 Disney films.

seuse and swordsman living in 1830s Japan. Here are montages of all 54 Disney films (1937–2014), 54 westerns (1939–2007), 42 gialli (1956–2013), and the 26 *Zatoichi* films (1962–89), each arranged in chronological order (figs. 7–10).

Having the corpora arranged in this way allows for a quick visual comparison, inviting researchers to consider what is common among the summed images as well as which individual images

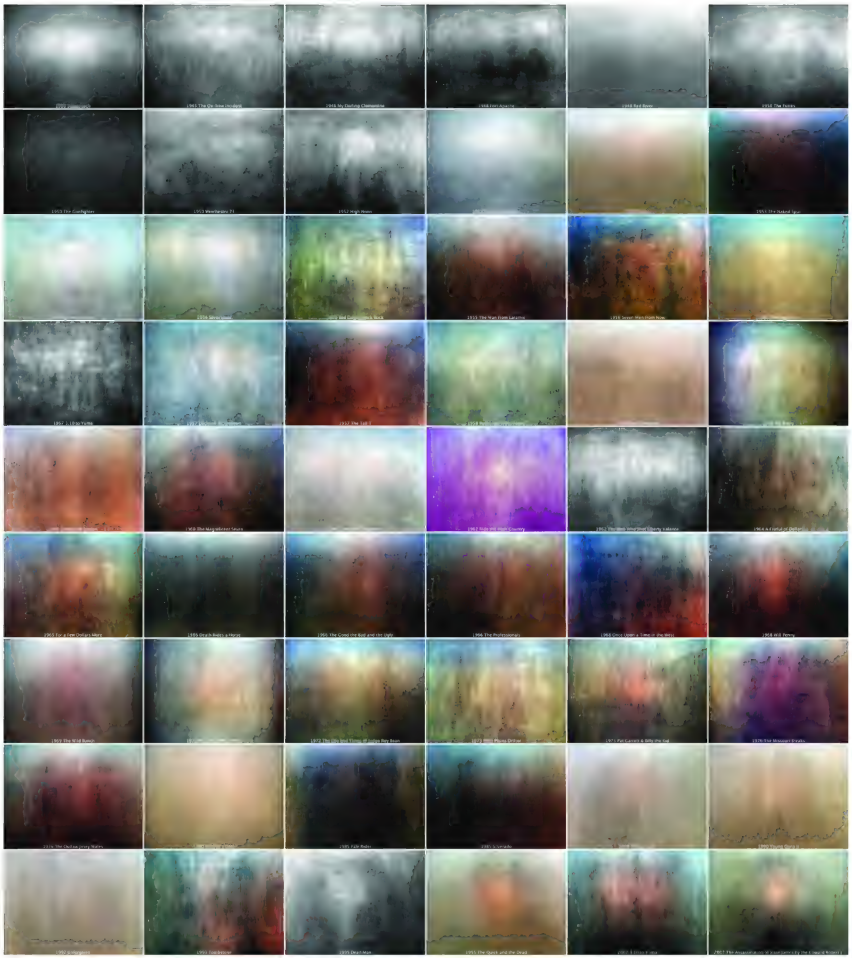


Figure 8. Summed frames of 54 western films.

stand out as unique. In doing so, we no doubt bring to bear our own assumptions about what we expect these corpora of summed films to look like: we might anticipate the Disney films to be more brightly colored than the gialli, for there to be some kind of subtle but discernible chronological change over the course of the *Zatoichi* franchise, or for the western to have a narrower range of visual difference. We might also be alternately dismayed and

Figure 9. Summed frames of 42 gialli.

surprised at our results, asking why particular films appear as outliers and investigating why some films do not appear as unique as we would have expected.

What is readily apparent upon first inspection is that these images do take a consistent form: they are all primarily composed of a lighter, centrally aligned shape of varying precision on a different-hued background with some degree of darker vignetting in the corners. Thus, whether a child-oriented animated film like *Bambi* (dirs. James Algar, Samuel Armstrong, David Hand, Graham Heid, Bill Roberts, Paul Satterfield, Norman Wright, 1942) or a violent live-action film like *Opera* (dir. Dario Argento, 1987), the summed film frame has a fairly narrow, impressionistic appearance. But

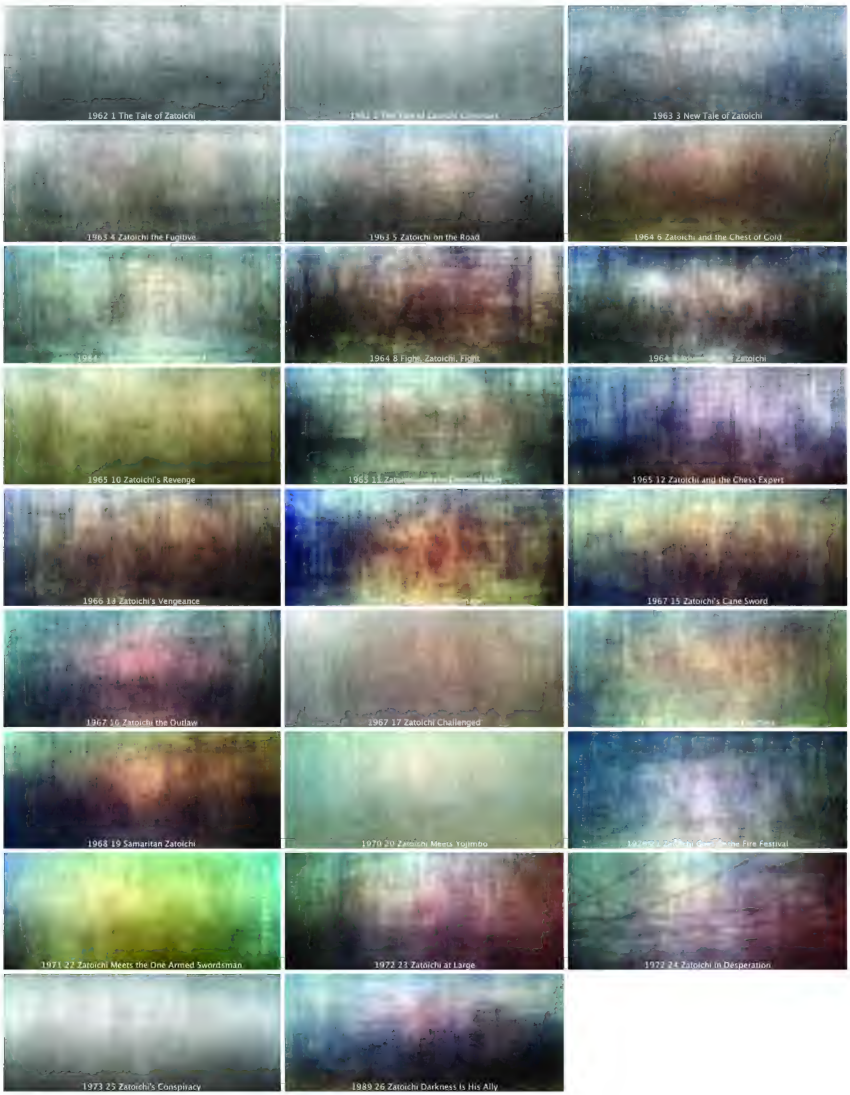

1962 1 The Tale of Zatoichi
1963 3 New Tale of Zatoichi
1963 4 Zatoichi the Fugitive
1963 5 Zatoichi on the Road
1964 6 Zatoichi and the Chest of Gold
1964 8 Fight, Zatoichi, Fight
1965 10 Zatoichi's Revenge
1965 12 Zatoichi and the Chess Expert
1966 14 Zatoichi's Vengeance
1967 15 Zatoichi's Cane Sword
1967 16 Zatoichi the Outlaw
1967 17 Zatoichi Challenged
1968 19 Samaritan Zatoichi
1970 20 Zatoichi Meets Yojimbo
1971 22 Zatoichi Meets the One Armed Swordsman
1972 23 Zatoichi at Large
1972 24 Zatoichi in Desperation
1973 25 Zatoichi's Conspiracy
1989 26 Zatoichi Darkness Is His Ally

Figure 10. Summed frames of 26 *Zatoichi* films.

the consistency in appearance also allows for examination of subtle and curious differences. Is there a reason why some films have strong vignetting and others little to none? Closer analysis shows that the ones without vignetting have more exterior shots

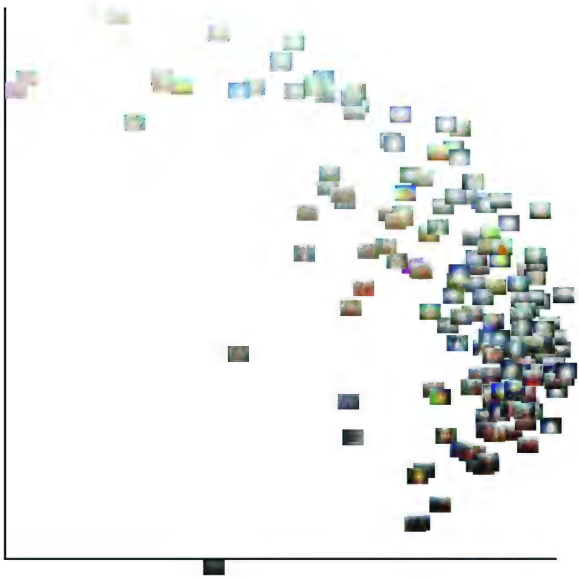


Figure 11. Entropy and brightness style space of the four corpora.

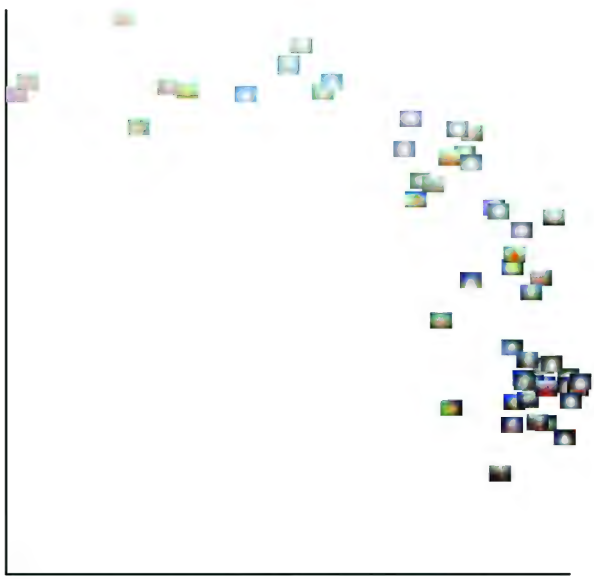


Figure 12. Entropy and brightness style space of the Disney corpus.

(and that they thus tend more likely to be sky blue hued at the top). Why do some films have the impression of a straight line down the center? This line is the space left between end credits, which appear as lighter blocks with a centered gap between role and name, and which only appear (and appear more prominently) in more recent films. Is there a reason why some films are so purple hued? That could be a subtle effect of the film stock source material, or a problem with overall color balance in the film transfer otherwise too subtle to notice. Visually scanning a corpora of summed film frames offers researchers an initial way to subjectively identify areas and questions of further interest. Ideally, this approach also defamiliarizes narrative film texts to prepare researchers "to see something as something other."

Once abstracted, the summed images themselves can then be measured and further compared. Giving shape to the initial subjective assessment, this second-order statistical measurement offers a way to more concretely compare the sum of a film's visual details. One way to do this is to compare corpora within a defined "style space," locating individual films in a two-dimensional space by plotting the relation of two chosen measurements, such as brightness, hue, entropy, or shape.[17] Here are five style-space graphs that do this. The first groups together all of the 176 Disney, western, giallo, and *Zatoichi* summed z-projections, plotting them by entropy (a measure of randomness, on the x-axis) and mean intensity (i.e., brightness, on the y-axis). This is followed by individual style-space plots showing the same graphs for each of the four corpora by themselves (figs. 11–15).[18]

Plotted this way, researchers can quickly see patterns within and between the four corpora, particularly noting places where clear groupings appear, offering suggestions for further investigation. First, the majority of the films demonstrate a similar level of entropy, clustering thickly in a rough column on the right of the x-axis, while displaying a much broader range of intensity along

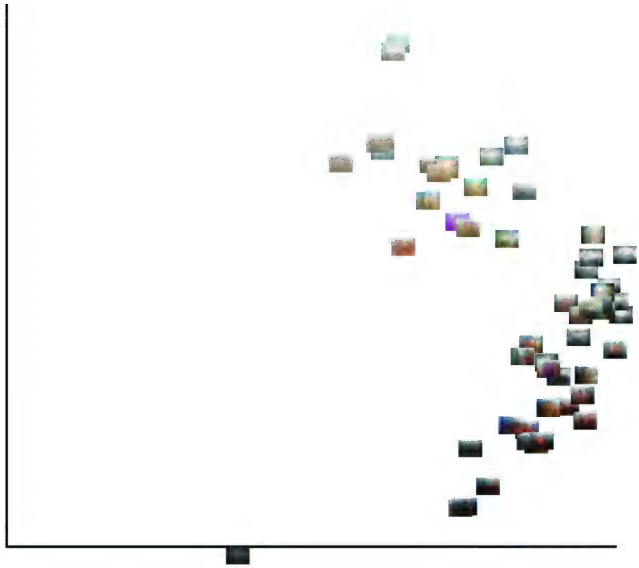


Figure 13. Entropy and brightness style space of the western corpus.

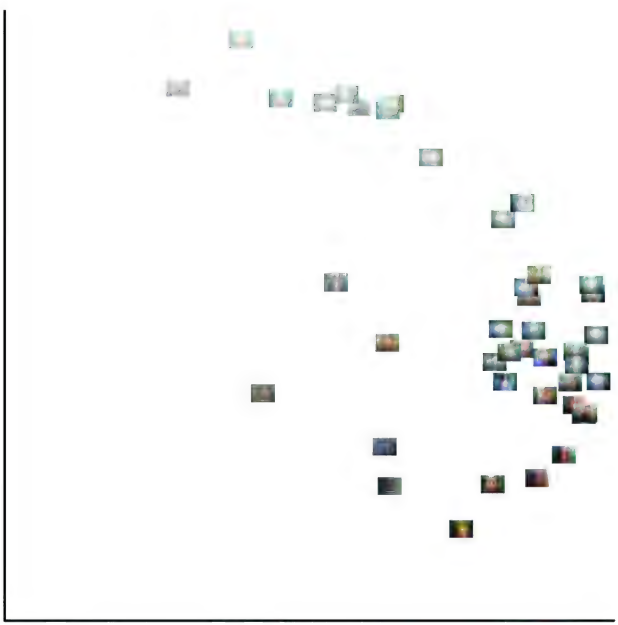


Figure 14. Entropy and brightness style space of the gialli corpus.

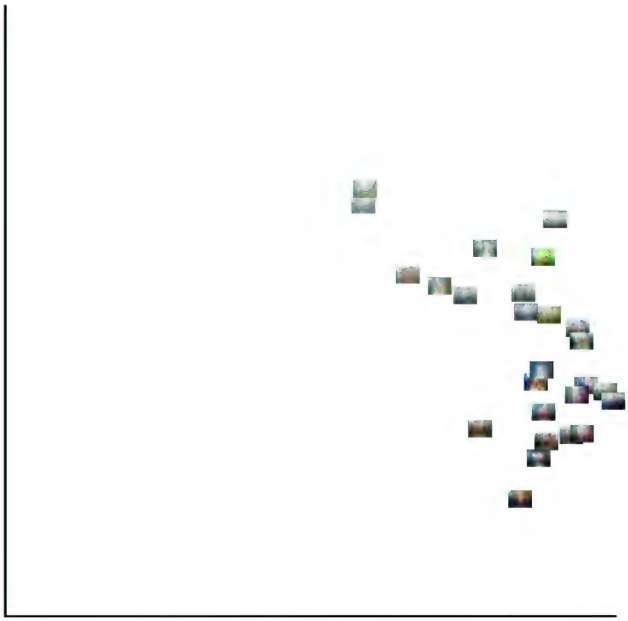


Figure 15. Entropy and brightness style space of the *Zatoichi* corpus.

the y-axis. There are, however, a significant number of similarly bright images at the top with varying entropy, and a number of other outliers that do not fit any larger shapes, such as *The Gunfighter*, sitting down alone on the bottom of the x-axis, which upon inspection is seen to be not only the darkest film of the bunch but also the least entropic of the western corpus. Indeed, looking more closely at the graphs for individual corpora, we see that with only a few outliers, the western is the most consistent, clustering neatly into two related shapes to the right of the graph; there appears to be a highly entropic, darker half-column and a second group of brighter, slightly less entropic films. In the middle of the latter group is *The Searchers* (dir. John Ford, 1956), which contrasts with a film like *The Missouri Breaks* (dir. Arthur Penn, 1976) from the first darker group. To account for the difference in these groups, researchers might consider a number of interpretive possibilities: are there more brightly lit exterior shots in films

like *The Searchers*, shot on location in Monument Valley? Or is it that there are more brightly lit interior shots filmed on sound stages? Would the difference between the groups also correlate to the year filmed, to the film process used, or to a division between types of westerns (the outlaw subgenre, spaghetti, or revisionist westerns)? Asking questions like these demonstrates how digital surrealism can produce interpretive lines of thought out of what was initially an explicitly experimental, irrational play with moving image materials. Rather than proceed from rational information that we already possess about film texts—their production year, credits, technical processes—we can first deform the films, adopt the resulting algorithmic unconscious as our own, and then work to make interpretive sense of this something as something other.

Having considered the difference within the western style space, we can also look to make comparisons between style spaces. We might initially notice that while the Disney corpus accounts for most of the brighter films, there are also nine gialli that occupy a similar space. Given that the content and intended audience for these two genres could not be more distinct, it is striking that they appear in part visually related. The gialli style space does not cluster as tightly as the western did, suggesting that while the genre is defined primarily by possessing a strong visual style, that style is not necessarily consistent. (And even with a smaller sample size, we can see clearly how the *Zatoichi* film sums occupy a strikingly similar space as the western corpus). But as with the western, we might ask if there is something else that the brighter gialli do have in common: a shared influence? country of origin? particular location? In doing so, we might discover unexpected visual affinities within and between radically different genres. Further investigation might compare other genres known for play with color and brightness, such as the film noir or the musical, to investigate not only comparisons between genres but also how individual films within a genre do or do not conform to expectations.

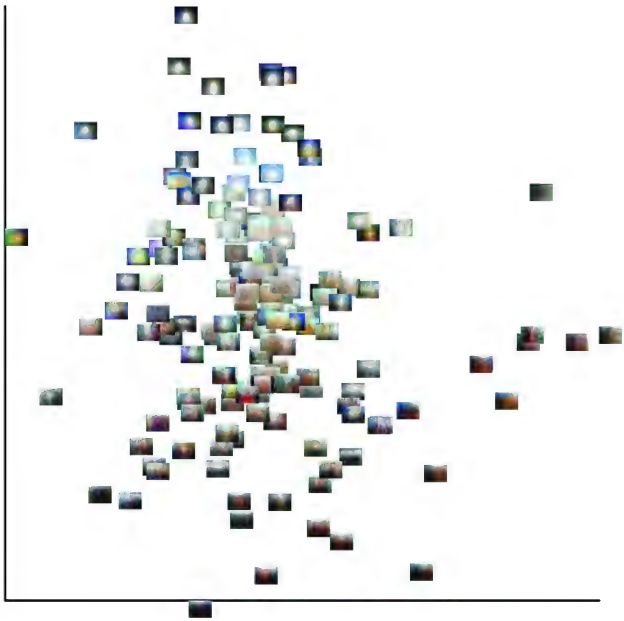


Figure 16. Plotting center of mass for the four corpora.

So far we have only looked at two visual features: brightness and entropy. But, if we want to think about another structural aspect such as composition, ImageJ can also measure the center of mass of an image, a "brightness-weighted average of the x and y coordinates [of] all pixels in the image."[19] We can use this approach to think about tendencies in framing. While I earlier described the summed frames as having a "centrally aligned shape," we see this is not strictly true when we look at all 176 films together (fig. 16), plotted by their center of mass.

First we note that the pattern of this graph is rather diffuse, with only the suggestion of a central vertical column. Setting aside the many outliers, we next notice in the central vertical shape a tendency to skew just left of the horizontal center, revealing an apparent preference for brightly-lit objects framed on the left of the exact center of the screen. This might accord with our expecta-

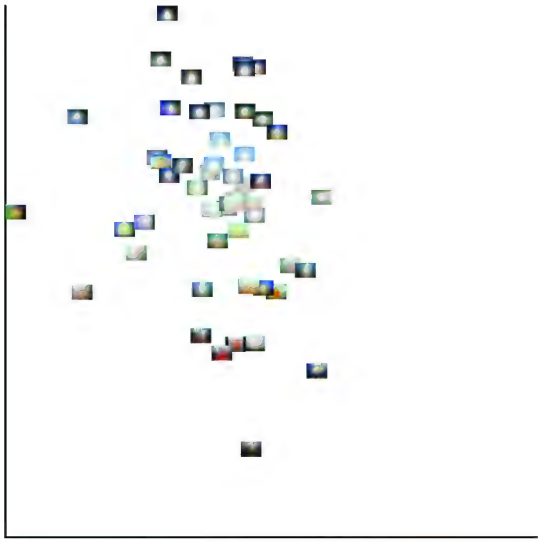


Figure 17. Plotting center of mass for the Disney corpus.

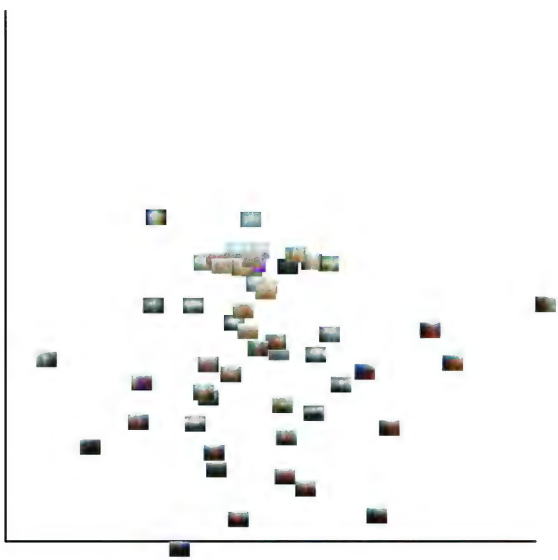


Figure 18. Plotting center of mass for the western corpus.

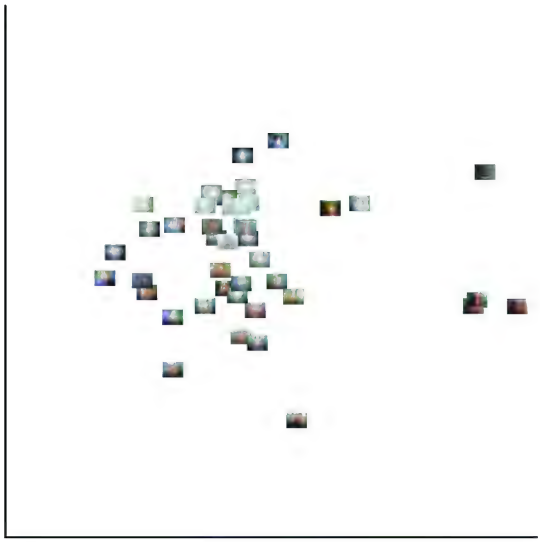


Figure 19. Plotting center of mass for the gialli corpus.

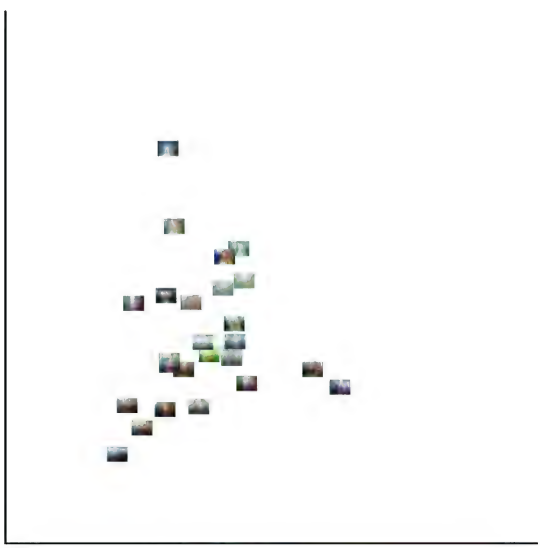


Figure 20. Plotting center of mass for the *Zatoichi* corpus.

tions for conventional ways of framing actors in dialogue to leave room for eye-line matches or aesthetic compositions, or the theory that Western spectators "read" a film from left to right as they do with text. Interestingly, though, there is no consistency in terms of a shared vertical alignment, with films stretched out all over the y-axis. Thus, while there are many exceptions, the sums reveal a preference for brightly lit objects framed on the left of the screen but with a range of vertical variance.

Above are plots of the center of mass of the four corpora individually (figs. 17–20: Disney, western, giallo, *Zatoichi*).

Looking at these corpora individually, we still note a range of variance, but it is clear the Disney and giallo corpora are much more consistent compared to the western corpus, which spreads over most of the graph. In fact, it appears as if, for the western, there is a preference for a center of mass below the middle of the frame (i.e., avoiding the top of the frame), whereas the other three genres align in various ways in a vertical column (i.e., mainly avoiding the right of the frame). This may in part be explainable by the use of widescreen aspect ratios (normalized in these plots for visual comparison), or it may suggest some otherwise invisible tendencies of lighting and framing (this would also be a question to ask of the four outlying, rightmost gialli). Whether it is a preference for asymmetrical framings, the effect of an emphasis on marginal objects, unusual lighting design, or some other factor is hard to tell from this graph alone, but it is striking that neither the Disney nor *Zatoichi* corpora approach this level of right skew. To that end, I am most interested in the *Zatoichi* corpus, which seems to argue in particular against the conventional wisdom that Eastern and Western spectators "read" film images starting from different directions. The *Zatoichi* films, with two exceptions, draw a strong vertical line indicating that their center of mass is consistently similar. Future investigation might further interrogate assumptions about the visual differences between Eastern and Wes-

Figure 21. Digital surreal manipulation of *Un Chien Andalou.*

tern films, comparing the center of mass of summed z-projections from other historical or national cinemas. By the center of mass metric, we can gain a sense of how a sampling of a range of films conforms to certain compositional patterns while some genres vary from the common trajectory. As such, the digital surrealist approach first creates irrational knowledge in order to provoke scholars into new lines of investigation into media history.

And the future? From the slice to 3D printing, anticipating a digital surrealist form of media historiography that finds surprising new points of comparison, the "little enough" when compared over a large corpus of films becomes something noticeably new. Embracing computer-aided, irrational, and automatic digital transformation as a research strategy offers a promise to reconceive media historiography from otherwise impossible to see vantages, slicing our vision so that we better see something as something other (fig. 21: *Un Chien Andalou* unsliced).

ENDNOTES

1 I gratefully acknowledge the assistance of Matthew Harrison, Robert Hunt, Charles R. Acland, and Eric Hoyt for their thoughtful contributions to an earlier draft of this chapter.

2 Tom Conley, *Film Hieroglyphs: Ruptures in Classical Cinema* (Minneapolis: University of Minnesota Press, 1991), xiii; and Robert B. Ray, *How a Film Theory Got Lost and Other Mysteries in Cultural Studies* (Bloomington: Indiana University Press, 2001), 2.

3 Johanna Drucker, "Humanistic Theory and Digital Scholarship," in *Debates in the Digital Humanities*, ed. Matt K. Gold (Minneapolis: University of Minnesota Press, 2012), accessed Nov. 8, 2015, http://dhdebates.gc.cuny.edu/debates/text/34.

4 Mary Ann Caws, *Surrealism* (London: Phaidon Press, 2004), 49.

5 Roland Barthes, "The Structuralist Activity," in *Critical Essays*, trans. Richard Howard (Evanston: Northwestern University Press, 1972), 214–15.

6 Barthes, "Structuralist," 215.

7 Franco Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History* (London: Verso, 2005), 53.

8 Ray, 100.

9 Adam Lowenstein, "The Surrealism of the Photographic Image: Bazin, Barthes, and the Digital Sweet Hereafter," *Cinema Journal* 46, no. 3 (2007): 62, accessed July 1, 2015, doi:10.1353/cj.2007.0024.

10 Roland Barthes, *Camera Lucida: Reflections on Photography*, trans. Richard Howard (New York: Hill and Wang, 1981), 55.

11 Paul Hammond, *The Shadow and Its Shadow: Surrealist Writings on the Cinema* (San Francisco: City Lights Books, 2000), 11–12.

12 Caws, 49.

13 Kevin L. Ferguson, "Volumetric Cinema," *[in]Transition: Journal of Videographic Film and Moving Image Studies* 2, no. 1 (2015), accessed July 1, 2015, http://mediacommons.futureofthebook.org/intransition/2015/03/10/volumetric-cinema.

14 Barthes, "Structuralist," 215.

15 Lisa Samuels and Jerome McGann, "Deformance and Interpretation," *New Literary History* 30, no. 1 (1999): 48.

16 Samuels and McGann, 36.

17 Lev Manovich uses the term "style space" to describe a visualization where "visual differences are translated into spatial distances" in order to be more easily compared. ("Style Space: How to compare image sets and follow their evolution," http://manovich.net/index.php/projects/style-space)

18 These graphs were created with the ImagePlot macro developed by the Software Studies Initiative, downloadable here: http://lab.softwarestudies.com/p/software-for-digital-humanities.html.

19 Tiago Ferreira and Wayne Rasband, "ImageJ User Guide, IJ1.46r," accessed July 1, 2015, http://rsbweb.nih.gov/ij/docs/guide/146-30.html#sub:Set-Measurements.

BIBLIOGRAPHY

Barthes, Roland. *Camera Lucida: Reflections on Photography*. Translated by Richard Howard. New York: Hill and Wang, 1981.

———. "The Structuralist Activity." In *Critical Essays*, translated by Richard Howard. Evanston: Northwestern University Press, 1972.

Caws, Mary Ann. *Surrealism*. London: Phaidon Press, 2004.

Conley, Tom. *Film Hieroglyphs: Ruptures in Classical Cinema*. Minneapolis: University of Minnesota Press, 1991.

Drucker, Johanna. "Humanistic Theory and Digital Scholarship." In *Debates in the Digital Humanities*. Edited by Matt K. Gold. Minneapolis: University of Minnesota Press, 2012. Accessed November 8, 2015. http://dhdebates.gc.cuny.edu/debates/text/34.

Ferguson, Kevin L. "Volumetric Cinema." *[in]Transition: Journal of Videographic Film and Moving Image Studies* 2, no. 1 (2015). Accessed July 1, 2015. http://mediacommons.futureofthebook.org/intransition/2015/03/10/volumetric-cinema.

Ferreira, Tiago and Wayne Rasband. "ImageJ User Guide, IJ1.46r." Accessed July 1, 2015. http://rsbweb.nih.gov/ij/docs/guide/146-30.html#sub:Set-Measurements.

Hammond, Paul. *The Shadow and Its Shadow: Surrealist Writings on the Cinema*. San Francisco: City Lights Books, 2000.

Lowenstein, Adam. "The Surrealism of the Photographic Image: Bazin,

Barthes, and the Digital Sweet Hereafter." *Cinema Journal* 46, no. 3 (2007): 54–82. Accessed July 1, 2015. doi:10.1353/cj.2007.0024.

Moretti, Franco. *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso, 2005.

Ray, Robert B. *How a Film Theory Got Lost and Other Mysteries in Cultural Studies*. Bloomington: Indiana University Press, 2001.

Samuels, Lisa, and Jerome McGann. "Deformance and Interpretation." *New Literary History* 30, no. 1 (1999): 25–56.

PART IV: **PROCESS, PRODUCT, AND PUBLICS**

DIGITAL TOOLS FOR TELEVISION HISTORIOGRAPHY: RESEARCHING AND WRITING THE HISTORY OF US DAYTIME SOAP OPERA

Elana Levine

The digital age has multiplied access to primary historical materials. In my work as an historian of American television, I have experienced a notable difference in available sources between the writing of my dissertation and first book in the early 2000s and a project I have been working on since 2008.[1] For one, there was no such thing as TV series released on DVD when I was first writing about 1970s television; syndicated reruns recorded to VHS were my source for those shows. In more recent years, however, other historians and I have benefitted from some institutional archives moving portions of their collections online, including full-text databases of popular, trade, and fan publications; online auction and second-hand sales services; and both officially distributed and user-generated streaming video sites. These points of access provide an array of primary materials that had been impossible or extremely difficult to find as recently as the turn of the twenty-first century. This wealth of materials has made possible my current book project on the history of the US daytime television soap opera, a foundational genre that has helped shape the medium and whose history sometimes leads and sometimes parallels that of broadcast network TV more generally.

Instead of enumerating the many primary sources I have found to help me build my archive for this project, this essay focuses on how I have been managing these sources by using digital tools. Due to the scope of my project (sixty-five years of television history), the voluminous nature of soap content (the genre airs new episodes daily fifty-two weeks a year), and the abundance of materials I have been able to collect, this project has presented challenges for storing, accessing, reviewing, and writing that I

had not encountered previously. In what follows, I detail my use of such tools as video downloading and conversion programs, scanning and optical character recognition applications, a data management system, and writing and mind-mapping software. In so doing, I argue for the distinct relationship between such tools and the practice of television historiography in the digital age, a relationship that has made possible that which was once deemed *im*possible. In 1985, US cultural historian Robert C. Allen noted the seemingly insurmountable challenges a history of soap operas might pose.[2] While his cautions abide in some respects, digital tools have helped to alter the parameters of possibility in exciting ways.

MANAGING VIDEO

In the integrated approach to television history I employ, TV programs themselves are but one site of inquiry. My research places programs within a cultural circuit that includes the industrial forces that produce television and the audiences that consume it, and that situates media within a broader historical context of social and political forces. Yet a project like mine, which historicizes a long-running TV genre, still sees the TV text as a significant primary source. The particular texts I am studying pose distinct challenges. Many were broadcast live and never recorded in any form. Even once the soaps were shot on video, the tapes were rarely preserved. And, for those programs that do have extant episodes, the volume of daily installments—between thirty and sixty minutes in length, broadcast five days a week, fifty-two weeks a year, for decades on end—makes it impossible to view most soaps in anywhere near their entirety.

Nevertheless, my strategy has been to watch what I can, particularly programs I have never seen in my own years as a viewer, either because they predated my soap-watching era (which began in the early 1980s) or because they fell outside my typical shows. Even within such guidelines, I am necessarily selective and my

choices are shaped by what I can access. In some cases, this means particular episodes preserved in institutional archives or uploaded by fans to user-generated streaming sites. While the former is a viewing mode long available to researchers, the latter is an access point only possible in the digital age.

In the years leading up to my beginning to work on this project in earnest—really since my teenage fandom of the 1980s, but also into the early 2000s—I began my own soap archive with historical or "special" episodes that appeared on television, saving them to VHS tape. Through the now-defunct SoapNet cable channel, I continued to save such materials, recording to DVDs instead of VHS tapes. As I committed to doing this project, I knew that I would want to integrate the viewing of my growing archive into my work routines. This is where digital tools began to play an important role. Had I needed to be tethered to a TV screen to play tapes or DVDs, or even to a computer with a DVD drive, it would have been difficult for me to make watching multiple soap episodes a day a part of my life. I would need to fit this viewing in around the multiple other work and life tasks that already filled my days.

My answer was to digitize this content and transfer it, first to a smartphone and later to a tablet. With both the content I have saved to DVD-Rs and content currently available via user-generated streaming sites, I convert or download to create MP4 files. To convert DVDs to digital files, I first used the software Visual Hub, which is no longer in operation, and then switched to Handbrake, an open-source software package. For content I access through user-generated streaming sites, I use downloading software, some of which is available for free online. I also use iSkysoft iTube Studio for its ability to download from a range of such sites, and to convert those files to MP4s. These steps mean that I spend some time nearly every day downloading, converting, and moving files. Luckily, much of this can happen in the background as I do other tasks. But having this content available to me on a tablet

that goes where I go, internet access or no, and can run alongside another screen is crucial to my workflow. I can fit in these episodes at many different times and places—on airplanes, while making dinner, answering emails, at the gym. I do take notes on what I watch, but often I am consuming so many episodes of a given program that only occasional moments are notable. Here it is important to remember that soaps are heavy on recapping and repeating content, that their visual styles are often formulaic, and that stories take many, many episodes to play out. Having watched enough of the genre, however, I have become expert in recognizing key narrative, visual, and sound examples when they happen. Thus when Dr. Steve Aldrich and Nurse Carolee Simpson start having multiple scenes together in the early 1969 episodes of *The Doctors* (NBC, 1963–82), I pay closer attention. I know these two will become one of the program's most beloved couples, and characters who had theretofore spent little time together having their own scenes signals me to the beginnings of their relationship. Much like the viewers who have made these programs part of their daily lives for decades, I am attuned to the relative significance of different soap moments.

My digital tools allow flexibility in watching soaps of the past that their original viewers did not have. Even those audiences from the time-shifting years of VCRs and then DVRs would not have had the mobility these tools allow me. Yet consuming the episodes in this way has become an unintentionally revealing experience. I've come to understand my viewing as the twenty-first-century digital version of the 1960s housewife glancing back and forth at the set as she irons, starts dinner, or moderates between squabbling siblings, an experience hilariously portrayed in a 1960 *TV Guide Awards* sketch. There may be no more fitting research strategy for a TV genre that has long served as a daily companion to its audience's lives. These digital tools not only give me access to these programs, then; they give me some sense of the original experiences of watching *and* listening to soaps, as my foremothers and I

have often relied on sound cues as we go about other tasks. That my consumption is less tethered to the domestic sphere than was theirs speaks to the changes in gendered social positioning my project also considers, as well as to the impact of digital tools on the research process.

STORING AND MANAGING DATA

As essential as digital tools have been to my ability to watch daytime soap episodes of the past, so too have digital workflows become the means of storing and managing the range of primary sources I draw upon in my research, from my notes on the soap episodes I watch to scans of scripts for episodes that no longer survive in moving image form. As in my archiving and reviewing of video, my digital practices for managing my materials more generally have developed over time.

For my first several years of research on this project, I employed the same methods I always had for gathering historical sources. I photocopied pages in manuscript archives and printed out articles from microfilm. I wasn't solely restricted to hard copies. I took notes in word-processing documents (though I did intend to print those eventually) and saved episodes I recorded off-air to DVD. By mid-2011, however, I began to realize that the scope of the project, and the volume of materials I was managing, would be much more usefully handled in virtual rather than physical form.

My previous workflow had been to work on a book-length project chapter by chapter. When I was ready to start structuring my ideas, I would sort my printouts and notes into piles. As I sat on the floor, these would accrue in stacks all around me, organized by topic or theme, until I saw patterns and trajectories coming together. This process worked well. But, due to my other commitments and the duration of my soap history project, I was researching many different portions of this new book at once, and planning to do the actual structuring and writing at a time that would,

potentially, be years in the future. Plus, I had so much material—even one chapter would likely mean many hundreds of pieces of paper surrounding me on the floor. Having increasingly digitized so many parts of my life and work in recent years, I decided to make my soap research project wholly digital as well.

I began to investigate tools for managing historical research materials digitally, settling eventually on a data management system called DEVONthink. I chose DEVONthink for a number of reasons: its compatibility with Apple computers, the ability to tag and add metadata, but mostly because it would allow me to perform optical character recognition (OCR) and make my many materials full-text searchable. Searchability was a crucial need, especially because I would be imposing a structure on my research after having built my archive over years and from multiple historical periods. It would be impossible for me to recall exactly what information I had about which topics; I needed to outsource that work to the software and be able to trust in its search functionality. This way, I could search a term like "divorce" to help me confirm a hunch that the 1958 soap *Today Is Ours* (NBC) was the first to feature a divorced heroine.

To make the database function, I had to digitize the paper archive I had already been building, turning paper into PDFs and importing word-processed notes as well. My ongoing archival research became about scanning rather than photocopying (using on-site scanners or a smartphone app, JotNot), generating more PDFs for the database. I began to use Adobe Acrobat to manage PDFs, combining pages into single documents when necessary, as well as doing some cropping and sizing to improve visibility. I have also used Acrobat to perform optical character recognition on large batches of recently digitized files, as well as running the OCR process in DEVONthink itself. Because I had some clerical assistance with the task of digitizing my paper archive, and my assistant did not have access to the DEVONthink software, using Acrobat was

the more efficient choice for those materials.

I imported these files into DEVONthink and I also began to generate all of my new notes within the software. Several years into using it as my data management system, I still have only a partial sense of its capabilities, as I have been figuring out how to fit it to my needs as I go.[3] There are many ways to tag or label or take notes on materials, including annotating and highlighting PDFs. The fact that most of my materials are searchable makes deliberately generating metadata less essential, although DEVONthink's ability to auto-classify documents improves with the more data one feeds it. I rely heavily on the highlighting feature to note key passages in materials that I might want to quote from or cite. And I've experimented with using the software's colored labeling system to help me keep track of which materials I have read and processed and which I have not.

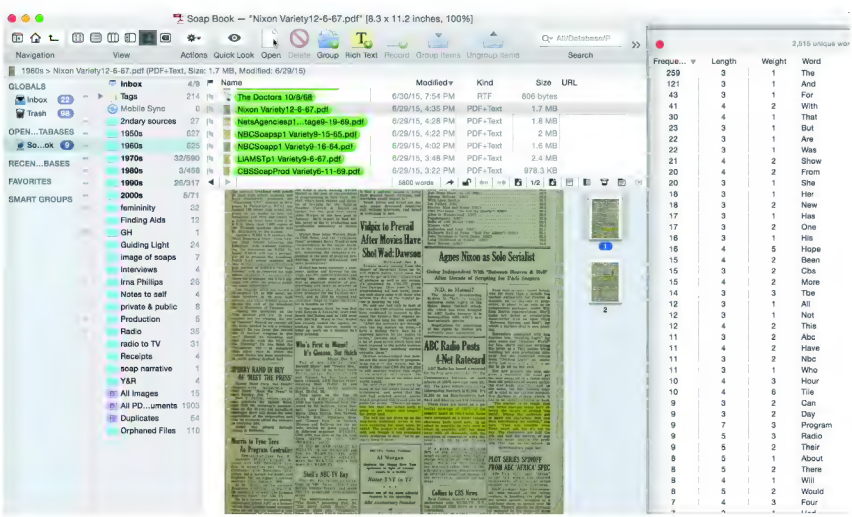


Figure 1. Screenshot from the author's DEVONthink database, demonstrating highlighting of PDFs, color coding of entries, and list of searchable terms from active document.

DEVONthink's optical character recognition is strong but OCR in general is an imperfect system, as blurry or nonstandard text may

not get appropriately recognized. Gradually, I have added tags and other notes to such files to make them more readily searchable later on.

Because I have figured out DEVONthink's utility as I've gone along, I've made some choices that I might make differently for another project. I initially put materials into folders (what DEVONthink calls "Groups") before realizing that was more processing labor than I needed to expend. So I settled for a sparser filing system, separating my materials into decades, but have taken advantage of a useful feature that "replicates" a file into multiple groups in order to make sure I put a piece of evidence that spans time periods into the various places I might want to consider it. For example, I might "replicate" a retrospective on the history of the soap, *Love of Life* (CBS, 1951–80), published in a fan magazine in the 1990s in my 1950s, 1960s, and 1970s folders, as the article includes details about each of those periods in the program's run. Had I invested more time in establishing more detailed folder hierarchies, I might get better use out of the software's ability to "file" documents automatically. I have settled into some file-naming practices, but would be more consistent about this on another go-round.

Still, the benefits of digitizing my materials in this system have been many. I have found that my ability to keep sources in multiple "piles"—whether through the groups I create or by using the software's search functions—allows me to make stronger and more nuanced historical arguments. I can put accounts of particular events into dialogue because I can easily bring materials from different manuscript collections together. For instance, I have been able to compare perspectives on the development of the half-hour soap in the mid-1950s by placing side by side sources from the trade press, from the unpublished autobiography of soap writer Irna Phillips, from the correspondence of sponsor Procter & Gamble, and from the files of agency/production company

Compton Advertising, all of which come from multiple databases, archives, and collections, gathered years apart. The facility with which I can move across and between sources, finding connections between far-flung materials, makes for more careful and detailed historiography.

In many respects, my workflow remains rather similar to my old, analog ways, in that I still spend long hours reading through all of my materials. But now I sort them into digital rather than physical piles, and can more readily allow materials to "fit" into multiple piles, representing different time periods or issues. The interpretive work of a cultural history of media does not change with these digital tools, and I am still the essential way station that determines what it all means and how it all fits together. I don't think there is a way to do that work without the time-consuming and pleasurable labor of reading and thinking, of sorting and categorizing, of articulating to each other that which a casual glance—or a metadata search—cannot on its own accomplish. But I have also come to believe that this particular project could not happen without digital tools. It is just too big and its details too numerous—how many soap storylines have you tried to keep straight at once?—to be feasible without the multifunctionality of digitization.

OUTLINING AND WRITING

This system of data management and storage was quite effective as I moved my archive into DEVONthink and continued to build it, adding materials as my research progressed. When I was ready to start writing, however, I realized that I had yet another digital need to fill. In my earlier, analog workflow, I would sort my paper research materials into piles, eventually labeling the piles with topics or themes on blank notecards that would sit atop their respective piles. With a legal pad and pen, I would sketch an outline of my chapter by figuring out the connections across the piles/categories, and testing out ideas for the big-picture arguments to

which the piles built. I had sorted my materials into decade-specific "piles" in DEVONthink. But I needed a much more detailed sorting of digitized sources to turn those random assortments of materials into chapters with structure.

For a while, I was resistant to considering writing software as the answer to this dilemma. Writing was not the problem. I had been writing digitally for a long time. Because I did so much planning and thinking before writing, I had no problem using conventional word-processing software to write. In fact, I like to write in linear fashion; it helps me construct a tight argument and narrate a coherent story. It was the outlining—the pile making, the planning and thinking—that I had to find a way to digitize. Then I saw the corkboard view for the writing software, Scrivener, which graphically reproduces the look of lined three-by-five-inch index cards. This virtualization of my physical piles made me reconsider my writing software aversion, and I decided to test Scrivener through its generous trial window (an option DEVONthink offers as well).

The trial sold me on the utility of the software for my process, although my use of it is quite specific and does not capitalize on all that Scrivener can do. Because I needed the software to help me to categorize my research materials and outline my chapters, I mainly use its "Binder" feature to sort my materials into digital piles. The hierarchical structuring of folders and documents within the Scrivener binder provides me with a way of replicating my mental and, formerly, physical labor of sorting and articulating ideas and information together in a digital space.

Having decided to organize my book chronologically, I began by reading through all of the materials in DEVONthink associated with the 1950s. As I read I categorized, figuring out what larger point the source spoke to or what circumstance it served as evidence of. I created what Scrivener calls "documents" for any piece of research or connected pieces of research that I thought

Figure 2. Screenshot from the author's Scrivener binder, demonstrating the outlining of chapters.

might be useful in my chapters. Early on, I realized I had multiple chapters to write about the 1950s, the crucial period when soaps transitioned from radio to TV and spoke to the postwar contexts of "containment culture" and the "feminine mystique," and ended up outlining three chapters at once as I moved through my materials. I gradually began to group documents into folders labeled

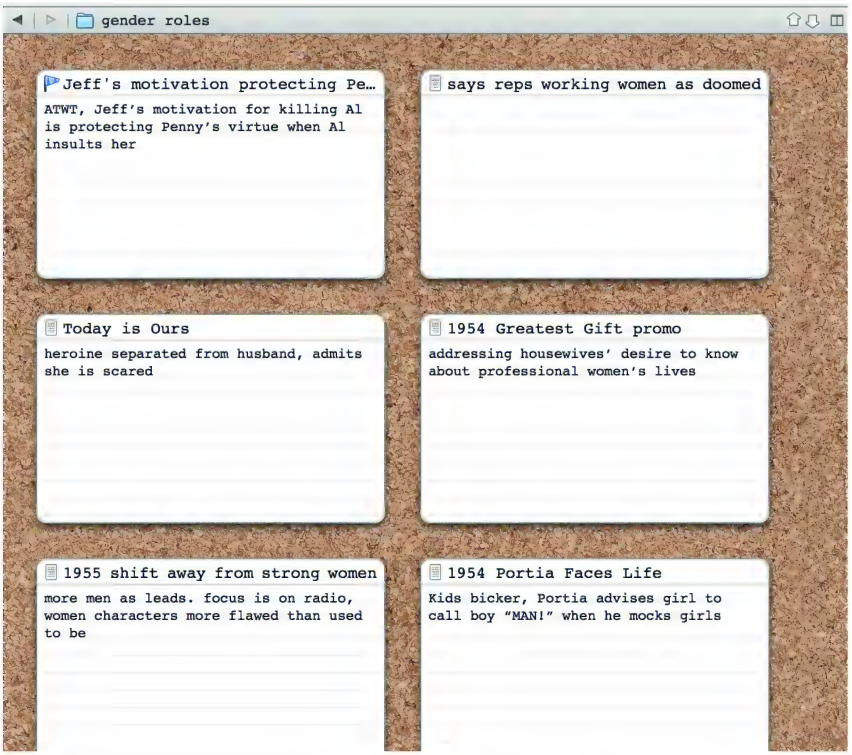

gender roles
Jeff's motivation protecting Pe…
ATWT, Jeff's motivation for killing Al is protecting Penny's virtue when Al insults her
says reps working women as doomed
Today is Ours
heroine separated from husband, admits she is scared
1954 Greatest Gift promo
addressing housewives' desire to know about professional women's lives
1955 shift away from strong women
more men as leads. focus is on radio, women characters more flawed than used to be
1954 Portia Faces Life
Kids bicker, Portia advises girl to call boy "MAN!" when he mocks girls

Figure 3. Corkboard view from the author's Scrivener outline.

with particular themes or points. This is the equivalent to me putting an index card with a label or category on top of a pile of papers, a way of understanding a set of specific pieces of information as contributing to a larger point or idea. These folders became subfolders of the larger chapter folders. But it is the way I integrate this process with DEVONthink that allows me to actually turn those digital piles into prose. In DEVONthink I am able to generate a link to a particular item in the database. I paste that link in the Scrivener documents I create.

How does this look in Scrivener? Sometimes this means that a Scrivener document is just my link, the text of which is the name of my DEVONthink item, such as, "SfT timeline late '50s/early

'60s," which are my notes on story events on *Search for Tomorrow* (CBS, 1951–82, NBC, 1982–86) during that period. But either within the document itself or in Scrivener's "Inspector" window, which can appear alongside the document on the screen, I can jot down notes about that source, reminding myself of the information it offers or indicating what I see as most relevant about it. The content I create here is what I see if I look at my documents in the corkboard view.

Other times my Scrivener documents include a number of DEVONthink links that feed into the same point. For example, a document called "Portia and Walter relationship" includes links to five different items in DEVONthink, four of which are notes on *Portia Faces Life* (CBS TV, 1954–55) scripts; the fifth is notes on memos from the show's ad agency producer to writer Mona Kent. In my synopsis notes on this document, I reminded myself that these were examples of the ways that married couple Portia and Walter talked to each other as equals, and how this served as a contrast to another couple on the show, Kathy and Bill. This ability to link to my DEVONthink archive has allowed Scrivener to serve as my categorizing and outlining system.

While I have written sentences here and there in Scrivener to help me remember the ideas I had about particular materials, I have not yet found need to actually write chapters within it. I use a conventional word-processing program for that. I know this is unlike the typical use of the software, but working this way has helped me to manage an otherwise unwieldy task. Scrivener provides a way to include research materials within its structure, but does not have the functionality of managing those materials that I get with DEVONthink.

This system is working well for me, but at times I do find the Scrivener binder structure to be too linear. The physical ability to move my paper piles around, to stack them or spread them apart

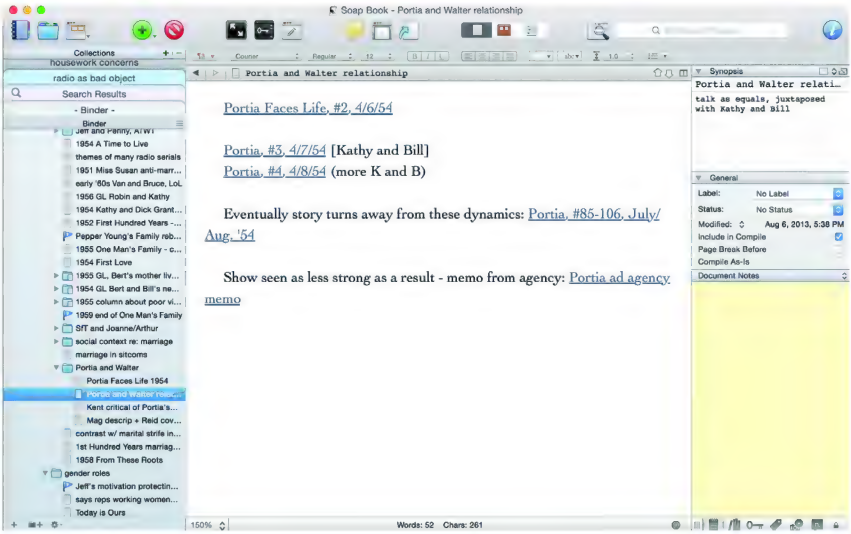

Soap Book - Portia and Walter relationship
Collections
housework concerns
radio as bad object
Search Results
- Binder -
Binder
1954 A Time to Live
themes of many radio serials
1951 Miss Susan anti-marr...
early '60s Van and Bruce, LoL
1956 GL Robin and Kathy
1954 Kathy and Dick Grant...
1952 First Hundred Years -...
Pepper Young's Family reb...
1955 One Man's Family - c...
1954 First Love
1955 GL, Bert's mother liv...
1954 GL Bert and Bill's ne...
1955 column about poor vi...
1959 end of One Man's Family
Biff and Joanne/Arthur
social context re: marriage
marriage in sitcoms
Portia and Walter
Portia Faces Life 1954
Portia and Walter relatio...
Kent critical of Portia's...
Mag descrip + Reid cov...
contrast w/ marital strife in...
1st Hundred Years marriag...
1958 From These Roots
gender roles
Jeff's motivation protectin...
says reps working women...
Today is Ours
Portia and Walter relationship
Portia Faces Life, #2, 4/6/54
Portia, #3, 4/7/54 [Kathy and Bill]
Portia, #4, 4/8/54 (more K and B)
Eventually story turns away from these dynamics: Portia, #85-106, July/Aug, '54
Show seen as less strong as a result - memo from agency: Portia ad agency memo
150%
Words: 52 Chars: 261
Synopsis
Portia and Walter relati...
talk as equals, juxtaposed with Kathy and Bill
General
Label: No Label
Status: No Status
Modified: Aug 6, 2013, 5:38 PM
Include in Compile
Page Break Before
Compile As-Is
Document Notes

Figure 4. Example of links to DEVONthink sources pasted into Scrivener document.

or move them in various ways was a helpful feature of my analog methods. As a result, I have begun experimenting with Scapple, a "free form text editor," similar to mind-mapping software and created by Scrivener's publishers, as a way to digitally reimagine the fluidity of the paper piles. Like Scrivener, Scapple allows me to link to DEVONthink items and has met my desire for a nonlinear planning system. I'm not convinced Scapple is essential to my digital workflow, but it does allow for thinking through materials in a different way.

Television historiography that understands the medium as functioning within a cultural circuit is always a somewhat unwieldy endeavor. Television's texts, running multiple seasons, can be many, many hours in length. The television industry involves an amalgamation of institutions—networks or channels, production companies, advertisers and their agencies, and government regulators—not to mention the range of individuals who fill key production positions. And television's audiences are impossible to differentiate from the public at large. How to isolate particular

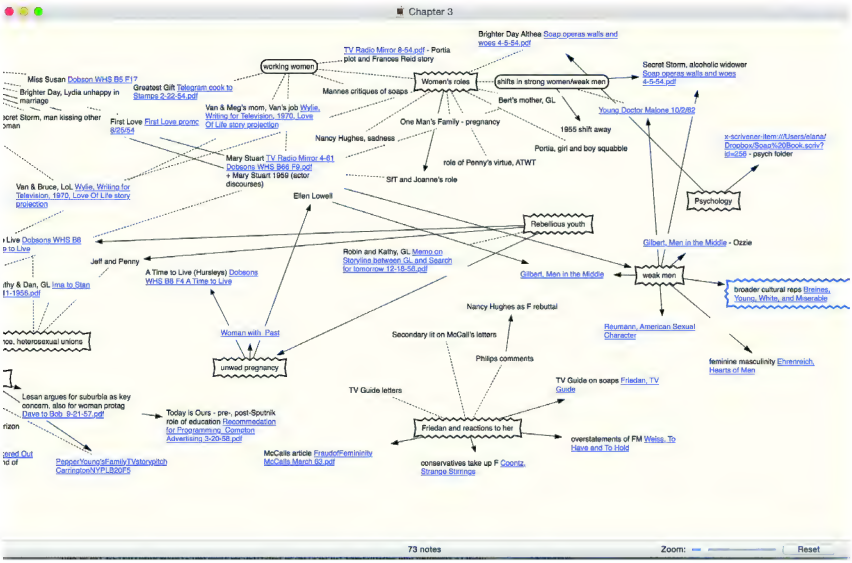

Chapter 3
working women
Women's roles
shifts in strong women/weak men
Psychology
Rebellious youth
weak men
unwed pregnancy
Friedan and reactions to her
Gilbert, Man in the Middle - Ozzie
Gilbert, Man in the Middle
Reumann, American Sexual Character
Nancy Hughes as F rebuttal
Philips comments
TV Guide on soaps Friedan, TV Guide
TV Guide letters
Secondary lit on McCall's letters
Woman wills, Past
Jeff and Penny
Ellen Lowell
Nancy Hughes, sadness
One Man's Family - pregnancy
Portia, girl and boy squabble
role of Penny's virtue, ATWT
SIT and Joanne's role
1955 shift away
Bert's mother, GL
Mannes critiques of soaps
Secret Storm, alcoholic widower
Young Doctor Malone 10/2/62
Miss Susan Dobson WHS B6 F17
Brighter Day, Lydia unhappy in marriage
Greatest Gift Telegram cook to Stamps 2-22-54.pdf
First Love First Love promo 9/25/54
Van & Meg's mom, Van's job Wylie, Writing for Television, 1975, Love Of Life story projection
Mary Stuart TV Radio Mirror 4-61 Dobsons WHS B66 F9.pdf + Mary Stuart 1959 (actor discourses)
Van & Bruce, Lol, Wylie, Writing for Television, 1975, Love Of Life story projection
Jeff and Penny
A Time to Live (Hursleys) Dobsons WHS B8 F4 A Time to Live
Robin and Kathy, GL Memo on Storyline between GL and Search for tomorrow 12-15-58.pdf
Lesan argues for suburbia as key concern, also for woman protag Dave to Bob 9-21-57.pdf
Today is Ours - pre-, post-Sputnik role of education Recommendation for Programming, Compton Advertising 3-20-58.pdf
McCalls article Fraud/of/Femininity McCalls March 62.pdf
conservatives take up F Coontz, Strange Stirrings
overstatements of FM Weiss, To Have and To Hold
feminine masculinity Ehrenreich, Hearts of Men
broader cultural reps Breines, Young, White, and Miserable
TV Radio Mirror 8-54.pdf - Portia plot and Frances Reid story
Brighter Day Althea Soap operas walls and woes 4-5-54.pdf
Secret Storm, alcoholic widower Soap operas walls and woes 4-5-54.pdf
x-scrivener-item://Users/elana/Dropbox/Scapb%20Book.scriv?itm256 - psych folder
73 notes
Zoom:
Reset

Figure 5. Mind-mapping ideas in Scapple, with links to sources stored in DEVONthink.

experiences or interpretations of television is an endlessly perplexing task. When I set out to study the path of the most voluminous of TV genres over the whole of the medium's history, I knew I would have to make careful selections of what to include and to exclude. But I also knew that there was a vast archive of relevant materials for me to draw from, and I wanted to design a project that could move nimbly across that multifarious terrain. It is a project that would require many years of my time and a commitment to a laborious research process. Over the time I have been working, however, I have come to realize that this is a project that also requires particular tools that digital technology has made possible. My book is very much mine. It is a product of my personal history as a soap viewer, my passion for archival research, my investment in relating micro-instances to macro-level developments, my memory for convoluted plots, my career stage, in which I can afford such an extended work process, and my commitment to work toward an end result located far in the future.

But the dependence of this work upon digital tools may be a sign of the direction media historiography can go with the assistance of this sort of technology. The story I am telling exists through my own efforts, but the tools that make it possible might shape television historiography in untold future directions.

ENDNOTES

1 See Elana Levine, *Wallowing in Sex: The New Sexual Culture of 1970s American Television* (Durham: Duke University Press, 2007).

2 Allen notes the voluminous hours of programming that constitute the history of even one daytime soap, as well as the fact that much of the genre's textual history has not been preserved. He characterizes a potential history of soap opera reception as "an enormous undertaking—and one fraught with any number of theoretical and logistical difficulties." Robert C. Allen, *Speaking of Soap Operas* (Chapel Hill: University of North Carolina Press, 1985), 13, 133.

3 Others have detailed their use of DEVONthink for historical projects. See Shane Landrum, "OCRing archival research photos with DEVONthink Pro Office," *Cliotropic*, last modified October 11, 2011, accessed March 1, 2016, http://cliotropic.org/blog/2011/10/ocring-archival-research-photos-with-DEVONthink/; Rachel Leow, "on DEVONthink and history research (I)," *A Historian's Craft*, last modified June 24, 2011, accessed March 1, 2016, https://idlethink.wordpress.com/2011/06/24/on-DEVONthink-and-history-research-i/; and Jamel Ostwald, "DEVONthink revisited," *Skulking in Holes and Corners*, last modified September 13, 2013, accessed March 1, 2016, https://jostwald.wordpress.com/2013/09/13/DEVONthink-revisited/.

BIBLIOGRAPHY

Allen, Robert C. *Speaking of Soap Operas.* Chapel Hill: University of North Carolina Press, 1985.

Landrum, Shane. "OCRing archival research photos with DEVONthink Pro Office," *Cliotropic*. Last modified October 11, 2011. Accessed March 1, 2016. http://cliotropic.org/blog/2011/10/ocring-archival-research-photos-with-DEVONthink/.

Leow, Rachel. "on DEVONthink and history research (I)," *A Historian's Craft*. Last modified June 24, 2011. Accessed March 1, 2016. https://idlethink.wordpress.com/2011/06/24/on-DEVONthink-and-history-research-i/.

Levine, Elana. *Wallowing in Sex: The New Sexual Culture of 1970s American Television.* Durham: Duke University Press, 2007.

Ostwald, Jamel. "DEVONthink revisited," *Skulking in Holes and Corners*. Last modified September 13, 2013. Accessed March 1, 2016. https://jostwald.wordpress.com/2013/09/13/DEVONthink-revisited/.

"TV Guide Award 1960 The Lost Art of Watching Soap Operas." YouTube video, 3:57, posted by "tvdays." Posted September 18, 2011, https://www.youtube.com/watch?v=6naYfqN73vM.

WHEN WORLDS COLLIDE: SHARING HISTORICAL ADVERTISING RESEARCH ON TUMBLR

Cynthia B. Meyers

"How good do that booty look though?" This comment was posted online a few months ago, in praise of an advertisement I had posted on my Tumblr blog (figure 1). The ad, from a 1946 issue of the *Saturday Evening Post*, consists of a Norman Rockwellesque painting of three children admiring Cellophane-wrapped lollipops. Above them is printed the question, "HOW GOOD DOES A LOLLIPOP LOOK?" and beneath them the answer, "YOU KNOW WHEN YOU *SEE* IT PROTECTED IN CELLOPHANE." What stood out for the anonymous online commentator, however, was the frilly-underwear-clad rear end of one of the children, a girl apparently about five years old, visible beneath her lifted skirt as she leans over the candy counter. Had I, in my effort to cater to my Tumblr audience, become the unintentional enabler of pedophiles? Or was this commentator simply a disinhibited social media user, anonymously poking fun at an historical image with an intentionally sexually perverse reading? How had I come to this moment of reckoning?

Initially, my intention in starting a blog was not to attract snarky comments but to share materials about our commercial cultural past. I study the history of the intersection of the advertising and broadcasting industries, a particularly rich topic, I think, in that both industries are deeply involved in creating, responding to, and disseminating a variety of discourses and cultural forms,[1] with sometimes distinct but often overlapping resources and purposes. The advertising industry itself is not a single entity but many institutions and individuals with conflicting and contradictory ideas and practices, pulled this way and that by competing assumptions and economic exigencies.[2] My book, *A Word from Our Sponsor: Admen, Advertising, and the Golden Age of Radio*, is the

HOW GOOD DOES A LOLLIPOP LOOK?
YOU KNOW WHEN YOU SEE IT PROTECTED IN CELLOPHANE
WHY IT PAYS TO BUY CELLOPHANE-PROTECTED PRODUCTS
DUPONT
Cellophane
SHOWS what it PROTECTS
BETTER THINGS FOR BETTER LIVING . . . THROUGH CHEMISTRY

"How good does a lollipop look?" (1946)

7 months ago 291 Share

Figure 1.

story of the role of advertisers and their agencies in broadcasting from its beginnings in commercial radio in the 1920s to the advent of television in the 1950s, a role which deeply affected, and continues to affect, not just broadcasting's institutional and economic structures but also its myriad cultural forms.[3]

This subject presents particular challenges to the cultural historian. Unlike print media, a large amount of which has been preserved, electronic media artifacts are scarce. Many live radio and television broadcasts were never recorded; of the recordings that were made, many have been lost or destroyed or otherwise made inaccessible.[4] So, like many other electronic media historians, I follow a paper trail of written documentation of advertising and broadcast industry practices—memos, correspondence, contracts, house organs (internal newsletters), publicity materials, trade publications,[5] and the like—that will expose framing assumptions and beliefs about audiences, entertainment, and advertising.

Memos and correspondence from "behind the scenes" are not likely to interest nonspecialist audiences, and the idea of sharing them publicly never occurred to me. However, as I shifted from collecting material by making hard photocopies to collecting shareable digital images instead, I began to pay more attention to the visuality of some of the artifacts. Two events in particular moved me to think the world might benefit if I posted some of my discoveries online. When I searched for images to serve as illustrations for my book as I prepared it for press, I was struck by how much information visual artifacts might supply readers seeking to understand the culture of the time. Advertisements, magazine articles, cartoons, comic strips, illustrations, and photographs vividly represent the role of radio in American culture. Print advertisements often cross-promote radio programs and their star performers. Broadcast and advertising trade publications feature advertisements that represent many of the debates and beliefs common in the industries. I found far too many interesting images

Penny and her mother both whistle this—
RIN-SO WHITE
EVEN LITTLE PENNY NOTICED HOW RINSO'S SOAPY-RICH SUDS GET OUT MORE DIRT
Rinso
AVOID SOAP WASTE

Penny and her mother both whistle this happy little washday song! (1944 Good Housekeeping)

8 months ago 24

Figure 2.

to fit in the book, and I wanted another outlet for them.

At about the same time, I was lucky enough to gain access to private archives that had been maintained for decades at one of the most important ad agencies of the period I study: BBDO.[6] In this archive I found folders, grouped by client, of several decades of magazine ads from consumer magazines such as *Life, Saturday Evening Post, Good Housekeeping*, and *Look*. These folders allowed me to review in succession ads for the same client from the 1930s through the 1960s and note how the art and copy strategies evolved over time; by posting their contents, I thought, I might draw online viewers into the same historical experience.

I noticed also that the agency would produce a prodigious number of ads based on a single concept or theme, often weekly, yet slightly alter the layout or illustrations or copy. Repeating the same slogan or concept was a basic hard-sell advertising strategy; variation might help prevent such repetition from boring the audience. For example, a 1944 ad shows "lovable little Penny Ann Vickers" and her mother whistling together to celebrate the fact that Rinso made little Penny's clothes clean as a whistle (fig. 2); a 1945 ad introduces "cute Patsy Anne Heinz" whistling "Rin-so White" in a black-and-white frame while her mother whistles "Rin-so Bright"

Rin-so White
Rin-so Bright
Why do you whistle on washday Mom?
Because Rinso safely GETS OUT DIRT—FAST from whites or colors
IT'S ANTI-SNEEZE
—the only granulated soap 98% free of sneezy soap-dust
Rinso
Why do you whistle on washday Mom? Because Rinso (1945)
1 year ago

Figure 3.

in an adjoining color frame (fig. 3); and a 1946 ad shows the eight-year-old Lally twins whistling both phrases in color as they bestride a stuffed horse (fig. 4). By showing my online viewers such series, I thought I might represent this common practice of repetition to them.

RIN-SO WHITE
RIN-SO BRIGHT
"WORK GOES TWICE AS FAST WITH RINSO'S TWIN ADVANTAGES"
GETS WHITES GLEAMING WHITE
KEEPS WASHABLE COLORS SPARKLING BRIGHT
Rinso
RINSO IS ANTI-SNEEZE
"Rinso is anti-sneeze-the only granulated soap 98% free of irritating sneezy soap-dust" (1946)
7 months ago

Figure 4.

In a sequence of Du Pont ads by BBDO for Cellophane, a word now in such common usage that most of us do not hear it as a brand name at all, I hoped to show how the agency's approach shifted

"How five sharp-eyed housewives got ideas for dinner tonight! Mrs. Nice, Mrs. Mater, Mrs. Phuss, Mrs. Clubby, and Mrs. Newly." (1939)

Figure 5.

"Today our metal is precious. And lots of use are serving Uncle Sam." "We're frosted foods in cellophane. We'll help to see you through." (1942)

Figure 6.

to accommodate the national mood during the Depression, the Second World War, and the postwar era. Cellophane evolves in these ads from an aid to thrifty housewives seeking to confirm the quality of their grocery purchases (fig. 5) to a patriotic household alternative to metals needed for the war effort (fig. 6) and a harbinger of futuristic technologies that will transform our lives (fig. 7). My readers would see memorably illustrated—through the ads' rich colors, elaborate layouts, and involved textual appeals—the process by which a large industrial company, Du Pont, was associated with the daily concerns of average consumers.

The intermediary between these pedagogical goals and my disinhibited commentator was the Tumblr platform on which I created the blog https://www.tumblr.com/blog/wordfromoursponsor.

Tumblr makes it easy to upload and share a variety of media (text, images, video, animated GIFs, audio, etc.). As Twitter developed as the dominant "microblogging" platform for text, Tumblr emerged as its first visual counterpart.[7] On a traditional blog, a user must navigate to the web page to see the post. On sharing platforms like Tumblr, users select other users' blogs to "follow" and then see the posts, in reverse chronological order, of every blog they follow without having to navigate from blog to blog. Like Twitter, Tumblr allows users to apply tags for easy searching. Tumblr's curators often highlight my posts with the hashtag "advertising" to promote their dissemination to the wider Tumblr community and the general public. My blog has just over 3,000 "followers" at this writing; it is one of several that specialize in "vintage" advertising.

"Mrs. Jones flicks a switch on her television set and tunes in the Shopping Tele-column of the Air." (1943 Du Pont ad)

Figure 7.

Like all scholars researching the past, I am careful to collect data that might help me understand the context of an artifact, such as its date of creation, original context, publisher, author, recipient, page, archive location, and so on. I try to understand the artifact's original audience and purpose. Ads from consumer magazines are obviously unlike those from trade magazines; publicity materials are quite different from internal memos or private correspondence. And they must be placed within the wider social, economic,

and cultural contexts of their production and reception to be properly understood.

Tumblr, I discovered, works to subvert all these aims and habits. Its users see my posts as decontextualized bits in a never-ending reverse chronological feed of all their followed Tumblr blogs, amid arbitrary adjacencies that prevent the building of meaningful contexts, even if a particular blogger tries to provide some. Some libraries and museums have created Tumblr accounts that try to counter these atomizing effects with lengthy commentary and explanation. For example, the National Archives posts thematically about certain historical topics, sharing documents and then encyclopedia-style entries about them. The Special Collections in Media and Culture based at the University of Maryland–College Park runs a Tumblr called @Bcast_Md that also provides encyclopedic information about its own posts, especially on the history of local broadcasting stations and specific programs. But these inevitably work against the tendencies of the medium. "The History of Flight" or "The Dedication of Mt. Rushmore" will appear contiguous, perhaps, to "The Worst Cat" or "Survivor: Beyoncé vs. Zombies," two of the most popular blogs of 2014 according to one site.[8]

Of course most social media users are looking at images for entertainment rather than for historical value, and, as historian/journalist Rebecca Onion notes, "The Internet loves a particular kind of history."[9] Certain subjects, eras, and visual elements garner more attention than others. As she selects which historical artifacts to post online at *Slate*'s The Vault, Onion notes, "[S]ometimes I'm like ugh, it's just too perfect. I can tell people are going to love it, but it's so pander-y that I almost can't." And catering to these proclivities may lead to oversimplification. As Onion explains, "What doesn't go viral: anything ambiguous, anything that doesn't tell a really direct story, that's not easily transmissible. I think the kind of history that does well on the web is the stuff

that's really unambiguous."

Figure 8.

My approach to sharing my historical research on Tumblr has, I fear, led me inevitably in the direction of the "pander-y" and the oversimplified. This process happened gradually. At first I had to learn how to build an audience by tagging and captioning effectively. I conscientiously tagged and captioned some of the basic metadata about the image, such as its year of publication, and highlighted in the captions elements, such as advertising slogans, that might interest users. I usually tagged the name of an advertiser that owned the brand (Lever Bros, for example, owned the Lux soap brand) and the advertising agency that produced the ad. I doubt any of this registered with my viewers. Most vintage-ad Tumblr blogs provide little or no metadata, and what is there may not be accurate; in fact, I sometimes see images tagged with the incorrect decade. My teenage daughter, then an enthusiastic Tumblr user, helped me to a clearer sense of what might attract actual viewers: she suggested I replace tags like "historical" with trendier equivalents like "retro" and "vintage." I also quickly learned that Tumblr users respond to tags that refer directly to visual qualities, preferably those of certain currently popular styles or eras, so I began including tags like "midcentury modern," "1950s style," "black and white" (fig. 8), and "illustration."

I found myself growing more and more interested not so much in

Who will be Miss WNBC-TV?

1 year ago 4

Figure 9.

GIFT
PREMIUM
CO.
Soggo
as impossible as trying to sell Portland without
KPTV channel 12
Best cost-per-thousand buy plus coverage and audience*—
that's why KPTV, Channel 12 is your MUST BUY media to sell the vast
Oregon and Southwest Washington market.
1 FEBRUARY 1958
61

The free premium of a young woman in a box of Soggo cereal is just as impossible as trying to sell Portland with KPTV-12 NBC. (1958)

1 year ago 194

Figure 10.

teaching the world as in gaining its attention. I learned the basic metrics Tumblr provides for each post: the number of "likes" (a heart icon) and "reblogs" (when another blogger reposts to his/her own followers). Tumblr does not show the number of "views" (although Google Analytics would do so) but lists the "likes" and "reblogs" as a combined list of "notes" and totals them. Like many social media users, I began to try to anticipate which posts would get the most attention. I began to select images not for their historical significance but because I thought they might attract more "notes." Often I was wrong. A silly photo of 1960s beauty queens vying for the title of "Miss WNBC-TV" got only four notes (fig. 9), while a 1958 trade publication ad featuring a sexy girl rising from

a box of Soggo cereal—her presence there as a premium is as "impossible" as selling to Portland without KPTV-12—got 194 notes (fig. 10).

I wanted my blog to consist almost entirely of content that was new to the internet, a quality that I thought at first would gain me followers eager for such material. Actually, however, reblogging is more important than originality on Tumblr, because reblogging others' posts is the best way to get them to follow your blog. The most popular blogs do this regularly and often, such as the Tumblr Klappersacks. Since I wasn't willing to automate my posts (line them up for automatic posting on a timed basis), post more than one new image per day, or reblog an image that didn't exactly fit my blog, I had to depend on rebloggers such as Klappersacks to spread my images for me; they became, in effect, my distributors, and therefore my primary audience. While I gain new followers almost daily, I believe most of them find my posts through rebloggers.[10]

What pleased these rebloggers? Celebrities. Sex. As a media historian, I prefer to study the unsung and unknown contributors to American commercial culture, who, given the collaborative nature of most commercial cultural production, are I think more important than a few auteurs or stars to the understanding of it. But when I noticed that movie star ads got more hits, I found myself posting more movie star ads. And from my extensive collection of Lux soap ads that featured movie stars from the 1940s through the 1960s, I found myself selecting ads with actresses still famous today (fig. 11). Stars sell, especially stars with currency.

The most reblogged of all my posts was an image of Mr. Spock and Captain Kirk gazing at each other from separate RCA televisions in a manner suggestive of longing (fig. 12). I benefitted, no doubt, from the size of the Star Trek fan community and the longstanding, half-joking supposition that these characters were more than

Judy Garland: "My Beauty Facials bring quick new loveliness" (1946)

Figure 11.

just friends.

Sexualized images, and, more surprisingly, sexist images were nearly as popular. Shamelessly I used tags such as “cheesecake,” “swimsuit,” or “sexism” to attract “likes” and “reblogs” (fig. 13).

See "Star Trek" on RCA Color TV
New, improved Automatic Fine Tuning. Just turn it on -- it automatically locks in correct fine tuning on all channels. The Mediterranean Ortega (left) and Italian Provincial Torino (right) feature 23" diag., 295-sq. in. color picture.
Spock & Kirk, together at last, on RCA televisions (1968)
1 year ago
488
Share
488 Notes
Hide

Figure 12.

And I found myself searching through my materials for ever more shockingly sexist images. At first I told myself that these images taught a valuable lesson about the bad past. But I began to wonder how they might help anyone actually understand their historical moment in any but the simplest, most reductive terms. Also I be-

gan to wonder what, after all, was their appeal? Was it the camp or kitsch that drew the hits? Was it actual genuine sexism? And then I got the comment with which I began this essay. A scene that for its original audience evoked childhood innocence and small-town wholesomeness inspired a current viewer to ask, "How good do that booty look though?"

Tumblr, I have been forced to concede, is not actually a place to learn very much about history, at least as a scholar frames and presents it. Most of my viewers probably do not care which ad agency created the ad or when or where the ad appeared or what the ad strategy was. Tumblr is a platform to share and circulate images. Rather than an historical archive, it gives us random juxtapositions and serendipitous discoveries. It lets us create and curate our own flow of images, original or reblogged. While Tumblr does allow easy search by tags, there is no way to reorganize posted materials—such as by date, topic, or source—so as to serve as an effective research tool. Furthermore, most of these images are so thoroughly decontextualized, both in how they appear in users' feeds and in the way most users tag and caption most posts, that Tumblr may seem to confirm the most pessimistic predictions for postmodernist culture.

I don't think it's quite this bad. Maybe the odd pedophile or sexist finds some accidental fodder among my images. Maybe most users would rather comment cleverly on perceived sexual allusions than consider how an image would have been received in its own era. Despite my efforts to provide context, I cannot impose historical understanding any more than I can prevent disinhibited online snark. What I prefer to think is that some users may, for a moment, look at the 1946 Cellophane "lollipop" ad, for example, and suddenly imagine a world in which such an ad is designed and distributed *without* the expectation that viewers would instantly sexualize a young girl's underwear. And a vivid realization of the differences of the past—different norms, expectations, and modes

talk about
IMPACT...!
ARB says it
"WLAC-TV has 41.0%
of the audience from sign on
The South's Great MULTI-MARKET Station
5
WLAC-TV
Talk about impact! (1958)
1 year ago
74
Share
74 Notes
Hide

Figure 13.

of reception—may place us briefly outside our own moment and help us understand it and ourselves: a venerable purpose of historical study. It's not history as I take such pains to shape it in my scholarly books and articles, where I can provide the contexts I think my artifacts properly demand, but it's history nonetheless, and I mean to continue to share it.

ENDNOTES

1 Jackson Lears, *Fables of Abundance: A Cultural History of Advertising in America* (New York: Basic Books, 1994).

2 John Thornton Caldwell, *Production Culture: Industrial Reflexivity and Critical Practice in Film and Television* (Durham: Duke University Press, 2008).

3 Cynthia B. Meyers, *A Word from Our Sponsor: Admen, Advertising, and the Golden Age of Radio* (New York: Fordham University Press, 2014).

4 Michele Hilmes, "Nailing Mercury: The Problem of Media Industry Historiography," in *Media Industries: History, Theory, and Method*, eds. Jennifer Holt and Alisa Perren (Malden: Wiley-Blackwell, 2009), 21–33.

5 Key trade publications include *Sponsor, Radio Showmanship, Printers' Ink, Advertising Agency and Advertising & Selling, Broadcast Advertising, Television Magazine, Tide,* and *Broadcasting*.

6 For a brief discussion of my research at BBDO, see Cynthia B. Meyers, "Using the Uncatalogued Archive," *In Media Res*, March 20, 2015, http://mediacommons.futureofthebook.org/imr/2015/03/20/using-uncatalogued-archive.

7 Instagram, a mobile rather than web application, has probably outstripped Tumblr as the primary image sharing platform as of this writing.

8 Madeline Stone, "25 Tumblrs That Went Ballistic in 2014," *Business Insider*, Dec. 3, 2014, accessed August 24, 2015, http://www.businessinsider.com/most-viral-tumblr-blogs-of-2014-2014-12?op=1.

9 Erin Loeb, "Beyond the Archives: An Interview with Rebecca Onion," *Vela*, accessed August 24, 2015, http://velamag.com/beyond-the-archives-an-interview-with-rebecca-onion/.

10 In the notes, Tumblr indicates which blog a user reblogs from, so it

is possible to see which bloggers are most influential by how many other users have reblogged from them.

BIBLIOGRAPHY

Caldwell, John Thornton. *Production Culture: Industrial Reflexivity and Critical Practice in Film and Television*. Durham: Duke University Press, 2008.

Hilmes, Michele. "Nailing Mercury: The Problem of Media Industry Historiography." In *Media Industries: History, Theory, and Method*, edited by Jennifer Holt and Alisa Perren, 21–33. Malden: Wiley-Blackwell, 2009.

Lears, Jackson. *Fables of Abundance: A Cultural History of Advertising in America*. New York: Basic Books, 1994.

Loeb, Erin. "Beyond the Archives: An Interview with Rebecca Onion." *Vela*. Accessed August 24, 2015. http://velamag.com/beyond-the-archives-an-interview-with-rebecca-onion/.

Meyers, Cynthia B. *A Word from Our Sponsor: Admen, Advertising, and the Golden Age of Radio*. New York: Fordham University Press, 2014.

Meyers, Cynthia B. "Using the Uncatalogued Archive." *In Media Res*, March 20, 2015. Accessed March 6, 2016. http://mediacommons.futureofthebook.org/imr/2015/03/20/using-uncatalogued-archive.

Stone, Madeline. "25 Tumblrs That Went Ballistic in 2014." *Business Insider*, Dec. 3, 2014. Accessed August 24, 2015. http://www.businessinsider.com/most-viral-tumblr-blogs-of-2014-2014-12?op=1.

NETWORKING MOVING IMAGE HISTORY: ARCHIVES, SCHOLARS, AND THE MEDIA ECOLOGY PROJECT

Mark Williams

The footage begins in medias res: dozens of African American women dressed in mourning attire have assembled in a public space and are organizing themselves into a peaceful, silent march around the square. We know it is a "silent" march because we see some women hush the others, and the scene is so quiet we can hear church bells begin to peal in the distance. Eventually it becomes clear that the women are marching around a mock-up of a tombstone in the center of the square, and that the location is Parker Center, administrative home of the Los Angeles police chief.

This example of historical television newsfilm is one of the key inspirations for the Media Ecology Project. It was among a selection of "raw" in-camera 1970s KTLA news footage rescued by the UCLA Film and Television Archive that was screened for the "Celebrating Orphan Films" conference in Los Angeles in May 2011.[1] Mark Quigley and Chris Horak of the UCLA Archive had invited me to select thirty minutes of clips from three hours of raw newsfilm to present at the conference. This was clearly the most historically important footage of the lot, even though I did not recognize the event or the name on the mock tombstone. Upon further research, it was identified as coverage of a 1979 demonstration in Los Angeles responding to months of inaction after the police shot and killed Eula Love, a recently widowed African American mother of three. Her killing was a turning point in contemporary civil rights debates in Los Angeles, though the memory of this event and its aftermath had largely receded from public memory. The footage may never have aired (this fact is not ascertainable), but the power and salience of the imagery is indelible and deeply instructive today regarding the value of historical newsfilm.

The Media Ecology Project (MEP) is a digital resource at Dartmouth College that enables researchers to digitally access archival moving image collections and footage, like the newsfilm described above. MEP contributes back to the archival and research communities through the fluid contribution of metadata and other knowledge. MEP seeks to enable new research capacities toward the critical understanding of historical media and to facilitate a dynamic context of research that develops in relation to its use over time by a wide range of users. The scope of MEP's work toward this goal includes exploring new methods of critical human and computational analysis of media, developing networks between institutions that expose existing archival collections to new audiences, and building tools that facilitate automated sharing of rich cultural data and metadata among software platforms.

My colleagues and I intend MEP to support and advocate the essential work of media archives, which range from the enormous holdings of the Library of Congress to Dartmouth's own media collections. Our moving image heritage is at enormous risk. Moving image archivists and digital repository advocates are developing solutions to the problem of preservation, but we cannot sustain interest in preservation without a better sense of the historical value of these materials. Access is not enough; new knowledge production is required in order to connect archival materials with audiences and prompt preservation and access efforts. MEP is working to produce cooperation and efficiency in relation to motivated engagement with academic communities.

The notion of ecology is central to the project in several ways. Those of us who work on media history recognize all too well that the materiality of historical media is fated. These historic materials simply will not endure without taking pains to preserve and archive them. In a fundamental sense this is a sustainability project: we are working to protect and ensure cultural and collective

memory in the form of historical media collections.

MEP networks media archives with new audiences; it also networks different software frameworks. The specific platforms we have engaged and are working to bridge are 1) Mediathread, a classroom platform developed at Columbia University that we are working to augment as a research platform; 2) Scalar, a digital publishing platform developed at the University of Southern California in relation to the Alliance for Networking Visual Culture; and 3) onomy.org, a new online tool that was developed for MEP and which will facilitate the creation of controlled vocabularies that can be assigned to online media files. For the time being, onomy.org is a stand-alone website, but this tool will soon be integrated with Mediathread and Scalar for MEP. The Media Ecology Project sits in between and in relation to these platforms and the participating archives, navigating the import, export, and production of metadata across participating archival content that has been engaged by a scholar or team of scholars. In this way we will contribute to the resultant capacities for search and discovery among these media elements in relation to others and realize new forms of research, scholarship, and publication.

PILOT PROJECTS OF THE MEDIA ECOLOGY PROJECT

The mission of MEP is realized through concrete results. We have four pilot projects currently in development that involve different archives and areas of media history. The common through line is the way in which they connect archives with researchers in order to achieve the shared goals of expanding access, generating new scholarship, and contributing valuable metadata for the archives.

Paper Print Collection: In conjunction with the Library of Congress we are engaged in a project regarding early silent film era materials, with an emphasis on the historically significant Paper Print collection, which is the equivalent of the Rosetta Stone for those who study moving image history in relation to visual cul-

ture. The Library of Congress has provided a first batch of one hundred Paper Print media files with related metadata for use in this pilot study and will continue to supply additional titles as the project proceeds. For this pilot study, we have enlisted Tami Williams (University of Wisconsin–Milwaukee) to codirect a core scholarly team consisting of several members of the renowned DOMITOR research society.[2]

In the Life: A second pilot study is focused on an important public television program, *In the Life*, which chronicled the history of gay and lesbian lived experience in the United States. The entire run of the program, plus all of the associated materials involved in its production (B-roll, interviews, etc.), will be digitized and placed online by the UCLA Film and Television Archive. We have begun to assemble a group of prominent scholars from the Society for Cinema and Media Studies to work on these materials, including Matthew Tinkcom (Georgetown University) and Stephen Tropiano (Ithaca College), plus members of the Gender Research Institute at Dartmouth. All episodes of *In the Life* have recently been uploaded to the MEP website, and the UCLA archive anticipates that additional programming materials will start to be digitized in 2016.

Historical News Media: The third pilot study connects the work of multiple archives and is dedicated to providing more and better access to historical news materials, including the 1979 silent march in Los Angeles. The archival materials include newsreels, news telecasts, newsfilm, and other associated footage. Archives who are participating include WGBH in Boston, the acclaimed bedrock of public broadcasting in the US, which features the groundbreaking Open Vault archive as an online resource and also hosts the Boston TV News Digital Library (news film materials from an array of local Boston television stations); the UCLA Film and Television Archive; the University of South Carolina's Moving Image Research Collections (MIRC), which includes the largest collection of Fox Movietone newsreels plus significant deposits of

local television station newsfilm; the Walter J. Brown Media Archives and Peabody Awards Collection at the University of Georgia, which holds multiple significant local television news film collections in addition to the unique historic programs found in the Peabody Awards nomination materials; Northeast Historic Film in Maine; affiliated online archival collections of television newsfilm at the University of Baltimore (deposited at the Internet Archive) and the University of Virginia; and the Library of Congress.[3] A core group of scholars has been assembled for this pilot project, including Mark Cooper (University of South Carolina) and Ross Melnick (University of California, Santa Barbara) who, along with Sara Beth Levavy (University of North Carolina) and myself, will coedit a collection of essays about newsfilm history in the United States. Francis Steen (UCLA) and Mark Turner (Case Western Reserve University) of the Red Hen Lab are leading the development of new research paradigms in relation to digital scholarship pursuits regarding the Newscape collection of television news broadcasts at the UCLA Library. Collectively, these networked archives and research projects have the opportunity to challenge and expand our assumptions about US broadcasting and American cultural history.

Films Division of India: This pilot is focused on studying the legacy of documentaries and informational films archived at Films Division in Mumbai, India. Since India's independence in 1947, Films Division has produced state-sponsored documentary, informational, and experimental cinema. This institution is also working to create a museum of national cinema that will curate and study the history of Indian cinema overall. An international team of scholars has begun the study of these materials.[4] One positive outcome that has already emerged from working directly with the Films Division is that our collaborative efforts have helped to reorganize their internal database about this historic collection. Whereas academia traditionally relegates the improvement of a database to a lower priority than publishing research books and

articles, we need to recognize that aiding and contributing back to media archives is a valuable form of scholarly output derived from our academic training and skill sets.

Toward this end, one of the goals we are pursuing in relation to each pilot study is the scholarly development of taxonomies or controlled vocabularies that can be deployed for the assignment of metadata to specific media content areas. The deployment and application of these vocabularies will enhance the functional discoverability of various archival content and augment future efforts to produce new forms of digital scholarship about these archival materials. MEP's archival connections are being built on public standards such as the Open Archive Initiative and the W3C Open Annotation format. Use of these widely available standards is key to realizing an ecology of applications that encourage bidirectional communication and share information as peers, treating archives as not just a source of raw materials but also as consumers of new analysis and scholarship.

BUILDING MEP

The Media Ecology Project was initially conceived in 2008 to fulfill a participant requirement at the foundational conference for new institutional members of the Mellon-funded Project Bamboo, an initiative that sought to develop shared tools and infrastructure for projects across the humanities. Most digital tools have been developed for the sciences, and there is a crying need for digital tools for the arts and humanities. As part of Project Bamboo, we recognized that it would be foolhardy to expect each institution to do everything necessary, and we must creatively think forward and collaboratively develop our goals with partner institutions. I gave a seven-minute presentation that identified a critical need for tools and infrastructure in the field of film and media studies, and identified the central conceits of MEP. When I brought up the goals of the project at the Association of Moving Image Archivists Conference later that year, I was glad to receive

immediate interested responses from the archival community. Progress on the Media Ecology Project has been greatly facilitated by internal funding sources at Dartmouth, including a Neukom Institute CompX Faculty Grant in 2012 to sponsor the initial building of the architecture of MEP and an enhanced capacity for annotation and tagging. John Bell was enlisted as the architect for the Media Ecology Project and has been an exemplary colleague and collaborator ever since. With funding from the Leslie Center for the Humanities at Dartmouth, we were able to convene an extremely productive symposium in May 2013, which brought together representatives from many of the MEP participating archives and institutions.[5] The symposium was successful in producing a series of agreements about the future of the project. One key initiative has been the development of a metadata server and attendant middleware that will mediate and build bridges between the Mediathread and Scalar platforms and will help to facilitate and maintain quality metadata produced in relation to archival elements. Bell designed and built the new open resource tool onomy.org in relation to the overall technical architecture for MEP, completing a triangle of tools and platforms that currently constitute MEP. The symposium generated significant interest across the archival community. MEP was featured at several conferences over the following two years.[6] The project received further support from Dartmouth to augment the development of metadata generation and capacities for curating annotations in relation to architecture enhancements derived from the development of our pilot projects.[7]

In 2015, Bell and I were awarded a two year NEH Research and Development grant to support the creation of the Semantic Annotation Tool (SAT), a drop-in module that facilitates the creation and sharing of time-based media annotations on the web. For the purposes of SAT, an annotation will consist of tags, a text body, and provenance metadata that describe a specific time-based or geometric fragment of a media file. This tool will be designed and

tested in conjunction with VEMI Lab (Virtual Environment and Multimodal Interaction) at the University of Maine. The finished tool will have two parts: a jQuery plugin that wraps an existing media player to provide an intuitive authoring and presentation environment for time-based video annotations; and a linked data-compliant annotation server that communicates with the plugin to collect and disseminate user-generated comments and tags using the W3C Open Annotation specification. Both parts will be released as open source software when they are complete. Potential uses include collaborative close reading of video for humanities research, simplified coding of time-based documentation in social science studies, enhancing impaired vision accessibility for media clips on websites, and many others.

We also seek to support the development of a scholarly-secure tier of access to online archives, an idea that all of our participating archives strongly encourage and support. Such a federated tier of access would open for consideration many collections that are restricted due to donor stipulations and ambiguous rights distinctions. This is an opportunity that will require a network of support and infrastructure to realize, and we are actively pursuing partners to help advocate and build such an infrastructure.

We are also actively developing relationships to varied approaches to machine vision applications. These tools have the capacity to produce large volumes of deep, granular metadata about small subsets of collections. We are keenly involved with the research of Dartmouth colleagues Michael Casey (the NEH-supported ACTION toolset which we codirect) and Lorenzo Torresani (Visual Learning Group) to enable computer vision and machine learning capabilities for moving image collections. We have worked with Torresani to develop an arrangement with the Internet Archive to extract a large benchmark dataset of historical news materials for use by academic researchers. Tools such as this will become essential for work on large deposits of newsfilm among other ma-

terials in our participating archives.[8] We are working to develop the MEP toolset to realize iterative efficiencies that support both manual annotation and machine annotation methodologies and lead to unique capacities for interpretation and scholarship.

LESSONS LEARNED

While developing a rather distinctive digital humanities (DH) project, we have learned first-hand several key lessons about this important and emerging field. Because we are building MEP from an arts and humanities perspective, we recognize that our goals must always be framed to raise awareness about the significance of cultural-critical perspectives within the various institutions that we have engaged (archives, libraries, universities, grant resources, etc.). Like many in DH, we underscore the need for collegiality and connectedness in pursuing collaborative work that depends upon openness and mutual respect as well as a balanced critical eye. Everyone who engages in MEP is at some level working outside their comfort zones: across disciplines, across expertise, across vocabularies. In a very real sense we are engaged in "translation" work, the great benefit of which can be experimentation regarding methodologies of study but also infrastructural designs of work-flow and output. We need to be vigilant about respecting difference and managing dissonance among highly skilled teams that often literally use the same words but do not speak the same language (e.g., metadata, annotation, ontology, research question, etc.).

Among the methodological comfort zones to be negotiated in digital humanities, several have become increasingly evident in MEP. We are committed to the development of visual culture studies in DH, which can produce tension with legacy approaches to DH that primarily focus on word culture alone. The field of film and media studies often features attention to research methods that address and engage audiences and the reception of media texts. Perhaps most importantly, and discussed further in Kit Hughes' chapter,

DH regularly features an iterative dialectic between the traditions of "close reading" in the arts and humanities versus the goals and practices of "distant reading" crucial to computational approaches to vast corpora of media texts under analysis. Recognizing these sites of potential dissonance will continue to be fundamental to progress in the emerging interdisciplinary space that is DH. MEP will continue to pursue twenty-first-century pedagogies and research procedures that contribute to the development of interdisciplinary approaches to visual literacy in relation to media history. In addition to extending the research profile of MEP as a networked resource, this will facilitate the widespread production of qualitative metadata that can support the essential work of the archives.

ENDNOTES

1 The Orphan Film Symposium is Dan Streible's extraordinary international movement that since 1999 has inspired many of us to action regarding film and media preservation and the rediscovery of forgotten media.

2 Participating scholars who have utilized the pilot study materials in their courses on silent cinema include Frank Kessler (Utrecht University), Laura Horak (Carleton University), and Amy Lawrence (Dartmouth College). The pilot study is coordinated into five Special Interest Groups who collaborate on examining varied aspects of the films in the collection. Presentations of their findings began in March 2015 at Columbia University.

3 We are especially looking forward to working with news and public affairs materials soon to be made available from the American Archive of Public Broadcasting, an extraordinary effort by WGBH and the Library of Congress to combine and collate all of the archives from public broadcasting stations across the United States. Much of this material will be radio and other audio files, for which our collaboration with the impressive and burgeoning Radio Preservation Task Force (also affiliated with the Library of Congress) will be deeply advantageous. But there will be an expansive television news and public affairs component to this col-

lection, both from local stations and from programs intended for national distribution, including the PBS NewsHour Collection of nearly 10,000 programs from 1975 to 2007.

4 We recently initiated a very promising research and scholarship effort with colleagues at the Academy of Film at Hong Kong Baptist University, including Camille Deprez, Ian Aitken, Emily Yeh, and digital and multimedia services librarian Rebekah Wong.

5 Details of the symposium can be found at "Media Ecology Project Symposium, May 17–18, 2013," Media Ecology Project blog, accessed March 20, 2016, http://sites.dartmouth.edu/mediaecology/symposium-may-2013/.

6 These conferences included: a dedicated panel about MEP at the Association of Moving Image Archivists Conference in Richmond, Virginia (November 2013), featured presentations on panels and workshops at the Society for Cinema and Media Studies Conference in Seattle (March 2014), as part of a panel presentation at an international meeting of the landmark Orphan Film Symposium in Amsterdam (April 2014), in a workshop on linked media tools and scholarship at the Extended Semantic Web Conference (ESWC) in Anassaris, Greece (June 2014), and as part of a featured workshop on digital tools in early cinema studies at the DOMITOR Conference in Chicago (July 2014), the PRELIDA workshop on Linked Data (Riva del Garda, 2014), EU Screen (Rome, 2014), Visible Evidence (New Delhi, 2014), Council on Library and Information Resources (Philadelphia, 2015), Project Arclight (Montreal, 2015), Open Repositories (Indianapolis, 2015), IAMHIST (Bloomington, 2015), and Moving Image Analytics (Stockholm, 2015).

7 Specifically, I am grateful to have received a second Neukom Institute grant in 2014 and a Scholarly Innovation and Advancement Award by the Dean of the Faculty at Dartmouth to facilitate my scholarship and travel in relation to the Media Ecology Project. I have also received a Mellon Leslie Center Medical Humanities Grant to initiate a possible research thread for MEP regarding the field of memory studies.

8 Additional computer vision MEP partners include Francis Steen and Mark Turner of Red Hen Labs, Virginia Kuhn (University of Southern California) who directs the Video Analysis Tableau, and Mark Boettcher in

Research Computing at Dartmouth.

CURATING, CODING, WRITING: EXPANDED FORMS OF SCHOLARLY PRODUCTION

Eric Hoyt

I have an elevator pitch for my digital humanities work that I've used for the past few years. It goes something like this: film and broadcasting historians have spent decades utilizing but not fully understanding the same handful of trade papers and fan magazines. I am engaged in a three-part process of 1) digitizing the diverse range of periodicals that we have previously ignored; 2) developing software tools to enable the search and analysis of those publications; and 3) writing books and articles that present the history of these magazines and model how media historians can use digital methods. As I always point out, none of this would be possible without the collaboration of an extraordinary group of scholars, archivists, collectors, and institutions.

There is one part of the pitch, though, that I've cut. I used to compare the way these projects worked together to a software suite, like the Adobe Creative Cloud. Beyond the mismatch of likening my open-access research to Adobe's proprietary software, the comparison gave a false impression of my workflow as harmonious. I made it sound like shifting across forms of scholarly production was as frictionless as moving a photograph within Adobe's Creative Suite from Bridge to Photoshop to Premiere. What I now realize is that these scholarly transitions are not seamless. Moving across these different forms of production is much more difficult and jerky than I once imagined. The workflows and processes are very different. The seams show.

In this essay, I reflect on those three forms of scholarly production that occupy most of my research time—building digital collections, developing software, and writing books and articles. In doing so, I attend to the ways their forms and workflows diverge

more than they converge. Even writing, which seems like the most conventional of the three scholarly forms, presents major challenges when digital methods become integrated into media historiography. Two important scholarly forms missing from this essay are blogging and online video producing, both of which I have engaged in on a limited scale.[1] I would encourage readers to explore MediaCommons and [In]Transition for models and reflections about the ways blogs, videos, and open peer review can be applied toward film and media history.

Although I share some of my own work and experiences, I hope this essay can be more than an exercise in navel gazing. As much as possible, I try to highlight the work of other scholars working within these forms and connect our efforts to broader theories and debates in the digital humanities. On a more practical level, I hope to give readers a sense of the resources available to them if they embark on these projects, as well as the challenges they are likely to face.

Perhaps most pressingly, this essay argues for the need to count digital collection building and software development as legitimate forms of scholarship. The question of what counts as scholarship is about more than tenure; it's about the ways graduate students are advised, the ways jobs are constructed, the prestige economy that nudges academics to take up one project instead of another, and the ripple effects of all of those decisions. Another flaw of my old software suite elevator pitch was its subtle suggestion of an implied priority. Magazines needed to be digitized so that the software could be developed, and I needed the digital collection and software so that I could finally proceed to the important work: writing! We need to move away from perceiving digitization and software as preconditions for the more significant work of analysis and writing. We also need to be wary about legitimating software and digital collections by suggesting that they can express the same arguments and do the same work as writing. They

don't. Instead, I believe we need to appreciate the contributions of digital collections and software on their own terms—terms as much about service to other scholars and a broader public as about expressing the developer's point of view.

1. DIGITAL COLLECTION BUILDING

Film and media historians today are actively engaged in digitizing historical materials and curating those artifacts. They are collaborating with archivists and librarians to make important films, radio broadcasts, and magazines openly available online. When intellectual property restrictions or a lack of extant copies prevents sharing the entire work, these scholars devise solutions, including sampling under fair use, writing the best possible descriptive text and metadata, and, in some cases, negotiating rights agreements. My own experience in building digital collections has been with the Media History Digital Library (MHDL). I codirect the MHDL with the project's founder, David Pierce, who, like me, wears dual hats as a film historian and digital curator.[2] The MHDL digitizes out-of-copyright books and magazines related to the histories of film, broadcasting, and recorded sound for open access. We have been able to achieve this through collaborating with institutions and collectors, who lend or give us the materials, and sponsors, who pay for the scanning.[3] The participation of the Packard Campus for Audiovisual Conservation at the Library of Congress has been especially transformative, enabling the MHDL's collections to double in size between 2013 and 2015.

Over the past five years, I have been involved at some point in every side of the MHDL's digital collection building process, ranging from the computer-based work of entering metadata and cropping images to the manual labor of packing up boxes of fragile magazines and shipping or hand delivering them for scanning (most, though not all, of the MHDL's scanning is carried out by the Internet Archive). David Pierce and I choose what to digitize based on a series of factors, including historical significance,

copyright status, physical availability, and input from our users and sponsors. David wisely started out by focusing on depth for a few key titles (e.g., lengthy runs of *Film Daily* and *Photoplay*) and breadth by having a single volume or two from a larger range of magazines (e.g., select years of *Shadowland* and *Film Spectator*). Because the collection has now grown to nearly two million pages, however, many lesser-known publications are now well represented. Our work is by no means finished, and we hope to greatly increase the MHDL's size over the next several years.

However, not all digital collections require the large size and scope of the MHDL. Indeed, if we evaluate collections only by size, then we risk overlooking many valuable resources and intimidating scholars from getting started on new collection-based projects. Table 1 lists over a dozen small-to-mid-sized collections that film and media scholars have had a hand in building. As the table reveals, the collections range from primarily credits and metadata (Canadian Educational, Sponsored, and Industrial Film Archive and Early Cinema History Online, described in more depth in chapters five and six) to archival document oriented (MPPDA Digital Archive) and moving image and sound collections (Phil Morton Memorial Archive, Jorge Prelorán Collection). The table could surely be much longer too. Additionally, there are ways for projects to productively collaborate. In 2016, the MHDL will be helping to provide access to a collection of late nineteenth-century international slide and magic lantern catalogs, curated and digitized by a team at Utrecht University.[4]

Still, there is much more work to be done in digitizing, curating, and enabling access to collections. And the work needs to begin by shifting our perception about what this work means and why it matters. Film and media historians often frame their involvement in these projects either as a precondition for research or an ancillary to a larger scholarly project. In other words, the scanning needs to happen so that the important analytical work can begin.

Name	Description	Website	Creators/Funding
Archivist of the "Yellow Peril"	A collection of representations of the "Yellow Peril" and early Asian Americans in Anglo-American print media, pamphlets, and newsprints. Includes movie posters, books, and advertisements.	http://www.apa.nyu.edu/gallery/kishi/	Yoshiro Kishi
Canadian Educational, Sponsored, and Industrial Film Project	This project focuses on the preservation and organization of industrial, educational, and sponsored Canadian films. Subjects include agriculture, the military, commerce, rural life, and many others.	http://www.screenculture.org/cesif/	Charles R. Acland with Louis Pelletier
Colonial Film: Moving Images of the British Empire	A website with information on over 6,000 films documenting life in the British Colonies, with more than 150 available for online viewing. Over 350 films have auxiliary critical notes written by their academic research team.	http://www.colonialfilm.org.uk/	The Colonial Film project is possible through the collaborative work of universities (Birkbeck and University College London) and archives (British Film Institute, Imperial War Museum, and the British Empire and Commonwealth Museum).
Columbia Screens	A website by and for students of film and media history at the University of South Carolina which explores how moviegoing evolved in Columbia, South Carolina, from 1904 to 1920. Using maps, newspapers, and other resources, Columbia Screens touches on how exhibitions intersected with daily lives, racial segregation, and urban design.	http://calliope.cse.sc.edu/colascreens/	Mark Cooper and powered by Omeka web publishing

Early Cinema History Online	ECHO is a filmographic database featuring credits for over 35,000 titles released in the US from 1908 to 1920.	http://echo.commarts.wisc.edu/	Derek Long with Paul Spehr, Susan Dalton, and Eric Hoyt
Going to the Show: Mapping Moviegoing in North Carolina	Going to the Show documents the experience of moviegoing in North Carolina from the introduction of projected motion pictures (1896) to the end of the silent film era (circa 1930). Employing maps, newspaper ads, photographs, city directories, and more, Going to the Show explores the intersections of moviegoing with race and urban and rural life.	http://docsouth.unc.edu/gtts/	Robert C. Allen with Natasha Smith, Elise Moore, Adrienne MacKay, Kevin Eckhardt, and Cliff Dyer. Going to the Show is made possible by the Institute of Museum and Library Services under the provisions of the Library Services and Technology Act as administered by the State Library of North Carolina.
Jorge Prelorán Collection at Human Studies Film Archives	A collection of the works of Argentine filmmaker Jorge Prelorán. Includes several of his films, audio recordings, production and correspondence files, and 31 digital books.	http://anthropology.si.edu/accessinganthropology/preloran/	Latino Initiatives Pool (administered by the Smithsonian Latino Center)

Name	Description	Website	Creators/Funding
Lost Films	Lost Films' mission is to collect and document film titles which are believed or have been declared as "lost." Lost Films is a platform where members can freely exchange, identify, and update information on these films.	https://www.lost-films.eu/	Deutsche Kinemathek – Museum für Film und Fernsehen with collaboration with: Bundesarchiv-Filmarchiv (Berlin), Friedrich-Wilhelm-Murnau-Stiftung (Wiesbaden), Centre national de la cinématographie (Paris), Filmarchiv Austria (Vienna), Národní filmový archiv (Prague). Lost Films has been made possible thanks to the generous support of the Kulturstiftung des Bundes (German Federal Cultural Foundation).
Margaret Herrick Library Digital Collections	A digital collection of the complete run of Academy Awards ceremony materials, including programs, posters, rule books, photographs, and other publications.	http://digitalcollections.oscars.org/	Margaret Herrick Library of the Academy of Motion Picture Arts and Sciences.
Media and the Movement: Civil Rights, Journalism, and Black Power in the American South	Media and the Movement aims to digitally preserve the media output of civil rights activists-journalists from the 1960s, 1970s, and 1980s. Documents include radio shows, interviews, and other private media collections.	http://mediaandthemovement.unc.edu/	Directed by Joshua Clark Davis and Seth Kotch with collaboration with Jerry Gershenhorn, Jacquelyn Dowd Hall, Joey Fink, Gordon Mantler, and Nicole Campbell. Supported by the National Endowment for the Humanities and the North Carolina Humanities Council.

Media History Digital Library	The MHDL digitizes classic media periodicals related to cinema, broadcasting, and sound. Collections include extensive runs of *Variety, The Film Daily, Modern Screen,* and many more.	http://mediahistoryproject.org/	David Pierce with Eric Hoyt
Medical Movies on the Web: Films from the National Library of Medicine	Medical Movies on the Web is a curated collection of medical films from the National Library of Medicine, from the Silent Era to the present. Films cover a wide range of health and medical-related topics, such as surgery, child development, diet, mental health, and much more.	https://www.nlm.nih.gov/hmd/collections/films/medicalmoviesontheweb/index.html	David Cantor, Michael Sappol, and Paul Theerman with the National Library of Medicine
Motion Picture Producers and Distributors of America Digital Archive	A database of the extant records of the General Correspondence files of the Motion Picture Producers and Distributors of America, Inc., from 1922 to 1939. Focuses on operations of the industry's trade associations and correspondences related to industry policies, public relations, censorship, and distributor-exhibitor relationships.	http://mppda.flinders.edu.au/	Richard Maltby and Ruth Vasey (with Jane Habner, Bruce Hatfield, Tim Cavanagh, and Liz Milford)
Moving Image Research Collections: Digital Video Repository	The MIRC-DVR is a wide-ranging digital repository for rare archival film and media materials. Videos come from all over the globe, including Chinese Films, US regional home movie collections, and science and nature films.	http://mirc.sc.edu/	University of South Carolina Libraries

Name	Description	Website	Creators/Funding
Northeast Historic Film	The NHF's mission is to collect and preserve film and video records of northern New England (Maine, New Hampshire, Vermont, and Massachusetts), and to provide public access to this history and culture of the region. Their collections contain 10 million feet of film and more than 8,000 hours of video; topics include amateur filmmakers, moving images of work life, and local TV newscasts.	http://oldfilm.org/	NHF has been awarded grants from New England organizations such as the Maine Community Foundation, the Betterment Fund, the Maine Humanities Council, the Davis Family Foundation, and agencies such as the National Endowment for the Humanities and the National Film Preservation Foundation.
Pare Lorentz Center at the Franklin D. Roosevelt Presidential Library	Established in the honor of award-winning documentary filmmaker Pare Lorentz, the center produces audio and visual materials to help teach history and social studies, modeled after Lorentz's social and political uses of the documentary format.	http://www.pare-lorentzcenter.org/	Elizabeth Meyer Lorentz (with a grant from the New York Community Trust to the Roosevelt Institute)
Phil Morton Memorial Research Archive	A free collection of audio-visual work by Phil Morton, a video artist and activist. A critic of current copyright laws, Morton's work was popular in the 1970s and was exhibited in New York (Museum of Modern Art), Chicago (Museum of Contemporary Art), and Brazil (São Paulo Art Biennial), as well as on major US television stations.	http://www.copy-itright.org/	Jon Cates with Barb Abramo. The archive is located in the Film, Video, and New Media Department at the School of the Art Institute of Chicago.

South Asian American Digital Archive (SAADA)	Documents and media related to South Asian American experiences. Includes materials on singer Kuldip Singh, radio and television shows, as well as media related to politics and community organizing.	https://www.saada.org/	Michelle Caswell and Samip Mallick
Texas Archive of the Moving Image	TAMI works to discover, preserve, and provide public access to Texas's film heritage. This online collection includes home movies, advertisements, local television, and amateur and industrial films as well as representations of Texas by Hollywood and international studios.	http://www.texasarchive.org/	Caroline Frick
The National Archives Unwritten Record Blog: Exploring History with the National Archives' Special Media Division	A blog by staff in the Special Media Archives Services Division, the Unwritten Record shares interesting encounters and discoveries during the archival process. These include media records such as analog and digital photographs, films, video, maps, and audio recordings.	http://unwrittenrecord.blogs.archives.gov/	The US National Archives' Special Media Division
Women Film Pioneers Project	The Women Film Pioneers Project (WFPP) is an online database that highlights the hundreds of women who worked behind the scenes in the silent film industry as directors, producers, editors, and more. Always expanding, the database features career profiles on each pioneer, longer overview essays on national cinemas and occupations, still and moving images, and archival and bibliographic resource materials.	https://wfpp.cdrs.columbia.edu/	Jane Gaines, Radha Vatsal, and Monica Dall'Asta, with support from Columbia University School of the Arts/Film and in partnership with the Center for Digital Research and Scholarship and Columbia University Libraries/Information Services.

Or, alternatively, the digitization occurs after all other research is complete, as a means of giving book readers the opportunity to go online and explore the primary sources for themselves. Both of these justifications can be true, and I have certainly been guilty of using them in the past. But I don't think we're giving ourselves enough credit—or pushing ourselves to do the best work possible—if we rely upon them.

Rather than framing digitization and curation as activities that are ultimately subordinate to other forms of scholarship, we should understand these practices as valid forms of scholarship in their own right. Literary scholar Jerome McGann has put forward one of the best cases for the matter, arguing for the renewed importance of philology, which studies how texts and languages change over time.[5] Philology's status within the humanities fell in the late nineteenth and early twentieth centuries as the work of creating critical editions of texts came to be perceived as less important than the work of interpreting texts. However, as McGann points out, philology's attention to how texts change across time and forms is valuable in our contemporary era of digitization.[6] And because these practices are valuable, we need to value them as a field.

McGann holds up the critical edition of a text as the definitive expression of philological practice; film and media scholars need to be more heterogeneous yet no less rigorous in our forms of collection building. The Society for Cinema and Media Studies would be an ideal organization to draft recommendations and best practices for scholars building digital collections and evaluating one another's digitization and curatorial work. This would help make it count toward tenure, but even more importantly it will encourage more scholars to get involved and create the best work possible.

Crucially, librarians and archivists need to be our allies and collaborators in this work. In many cases, the very labels and distinc-

tions of "scholar," "librarian," "archivist," and so on are unhelpful. Many of us have overlapping levels of expertise and all of us have ways to contribute that go beyond our titles. And the so-called amateur may possess a deeper knowledge about a particular topic than the professional. However, for those of us trained primarily as researchers, writers, and teachers, we should listen especially carefully to our librarian and archivist friends when they bring up questions of usability, findability, and preservation. No one wants to pour her energy into a project that researchers and the public never discover or want to use—or into one in which the data corrupts or disappears from the web.

2. DEVELOPING SOFTWARE

When film and media scholars build digital collections, they typically do so using a software platform, such as Omeka or WordPress, both of which offer graphical user interfaces (GUI). But what happens when the media scholar wants to build a digital project that an out-of-the-box software package cannot offer? She may find herself stepping into the waters of software development and design.

One could easily exaggerate the difference between developing software and building a collection-oriented digital project. In both cases, you are likely to use some of the same open source technologies, including relational databases (e.g., MySQL), search indexes (e.g., Solr), and coding languages (e.g., PHP, Javascript, CSS, HTML). But developing software—whether it's a tool like Arclight or a databased digital project like those described by Miriam Posner in chapter eight—requires that you actually understand how these technologies work on a much deeper level. To accomplish your goals, you may have no choice other than to rewrite lines of code and program entire new sections of an application. You will have to say goodbye to the comforts of a GUI as you enter prompts on the command line.

You can and should seek out collaborators to help with software development. Lantern and Arclight would never have been possible without the contributions of Carl Hagenmaier, Wendy Hagenmaier, Andy Myers, Pete Sengstock, Kevin Ponto, and Alex Peer, all of whom had expertise in programming languages that the other team members and I lacked. Nevertheless, I found that effectively leading these projects required that I develop a basic understanding of computer programming and open source technologies. Without these skills and knowledge, I would not have been able to communicate with the other members of my team. I would have also wasted a lot of their time by asking them to reinvent the wheel rather than adapting open source software packages that I had identified. And in the case of Lantern, there were moments when I was a team of one. I had to solve a programming problem because, if I didn't, no one else was going to do it. The project would have died. *For Dummies* guides, Lynda.com tutorials, and Googling error messages got me through many jams. I highly recommend all of these resources for any media historian considering trying his or her hand at software development. Most importantly of all, you will need a great deal of curiosity and patience.[7]

Of all the forms of scholarship that I've produced, I have found software development to be both the most exhilarating and frustrating. In the case of Lantern, the highs are easy to remember: the ah-ha of figuring out Ruby on Rails' model-view-controller architecture; the thrill of witnessing that a new algorithm, which took a year to develop, vastly accelerated the search speed; and most of all, the rush of publicly announcing Lantern's launch in the summer of 2013 and hearing the immediate positive feedback. But the frustrating moments are memorable too, and they were all too frequent. Software breaks down. I can't count the number of days I thought I would spend writing that were hijacked due to technical glitches. Even when things are working fine, they could always be working better; Lantern contains some broken links

and inaccurate metadata that users bring to my attention and I need to change. When the MHDL scans more magazines, it means that I need to index those magazines into Lantern so that they become searchable. This is a more complicated process than one might imagine due to how the MHDL's collections are organized on the Internet Archive and the toll that all of this takes on our customized version of the Apache Solr search index, which we have begun to outgrow. All of these updates take time but don't fit neatly onto any of the lines on my CV.

Ultimately, the time and headaches spent on maintenance are worthwhile because I know that thousands of people use Lantern and depend upon it for their research (our Google Analytics show that between three and five hundred users visit per day, with the average user session lasting around eleven minutes). I hope humanities tenure committees reach a point where they accept that building software and digital collections can contribute something as valuable, if not more, than a book or series of journal articles. Some digital humanists have made the case for software as a legitimate form of scholarship by suggesting the form, like a piece of writing, can be a vehicle for an argument.[8] The work of Kim Christen in developing digital archives for indigenous peoples meets this standard. Her work serves marginalized communities, but it also makes an argument that existing content management systems have been built with Western assumptions about technology and access. Digital humanities scholars have also made the case for "glitch art" as an argument—or at least intervention—that disrupts our familiar, noncritical interactions with software and exposes aspects of a technology that typically stay hidden.[9]

In my own work, though, I have found most of the arguments I've tried to embed into software have been failures—either the software doesn't work, the argument is overly simplistic, or the audience misses the intended argument.[10] When Anne Friedberg created an interactive digital project to accompany her book *The*

Virtual Window, she found that "the digital format is not at its best in building a complex argument; it works by accretion, by juxtaposition, by comparative assemblage."[11] Software can be highly suggestive, like a form of visual art, but it lacks the expressive clarity, precision, and linearity that most complex arguments require. After spending five years working in this space, I have come to believe that the best reason to develop software is not to advance our own arguments. Instead, we build software to serve others, allowing them to arrive at their own insights, surprises, and arguments.

3. WRITING

Writing remains the best form for constructing and expressing the sorts of complex arguments that Friedberg describes. Here, I'm conceiving of writing as a form that may include supporting illustrations, data tables, or media clips, but that is first and foremost driven by words, sentences, text. The presentation of the writing may be a peer-reviewed journal article, open access PDF, series of HTML documents constructed in Scalar or WordPress, university press paperback, or another format entirely. I have found, however, that thinking about genre is more useful than focusing on publication format. My writings and publications that sit at the intersection of media history and the digital humanities fall into one of three genres: the self-reflective essay, the essay-report, and the book-length monograph. I will try to briefly sketch out each of these genres and what I see as their strengths and limitations.

Most of the essays collected in this book are self-reflective. The authors share some of their projects and research, but this largely serves as a means for reflecting about their process and the broader implications for scholarship. Readers page through these essays less to see evidence and theory mounted to support an original historical argument and more to better understand the various digital humanities methods at play and to consider what they mean for media history. Reflective essays can be more per-

sonal and conversational than traditional research publications—more like the brief "In Focus" pieces at the back of *Cinema Journal* rather than the polished ten-thousand-word research articles that precede them. As a writer of a reflective essay, you are permitted, even expected, to raise questions that go unanswered.

One limitation of a reflective essay, such as the one you are reading right now, is that they can become quite insular. We are analyzing ourselves, rather than turning our analysis to questions that go beyond the traditions and idiosyncrasies of academic disciplines. In some cases, a reflective essay will transition into something more akin to a manifesto or position paper, arguing for a particular way forward based upon those reflections. But even reflective manifestos call for the innovation of new methods and theories (with titles that frequently begin "Toward a . . .") far more than they model what results such new theories and methods actually yield.

A second genre that retains an interest in methodology but seeks to be less personal and more results-oriented is the essay-report.[12] Willard McCarty, who coined the term, explains that the essay-report "draws on both the conventional essay in the humanities and on the laboratory report in the sciences." The essay-report's basic six-part structure is modeled on scientific papers:

* Introduction
* Method
* Results
* Discussion
* Conclusion
* References

In my own work, my collaborators and I have deviated from this structure somewhat—incorporating case studies and topic-specific sections, for instance—but we have always included the lengthy

Method section that distinguishes our essay-reports from other film- and media-studies articles we've written.

Another aspect of the essay-report that is more like the sciences than humanities is that these journal articles are frequently coauthored. The five essay-reports that I've published—four of which involve Project Arclight in some manner—have all been coauthored with at least two other people.[13] The culture of publishing coauthored articles (generally listed in order of involvement or, if all contributions are equal, alphabetically) acknowledges that most software development projects and large-scale data analysis experiments require the input of multiple investigators. The formulaic structure is also an advantage when it comes to joint authorship. After the team's data analysis is complete (or close to complete), the principal investigator can quickly divide the writing labor—ok, you write the Method section, you write up the Results, I'll handle the Discussion, and then we can all circle back to the Introduction and Conclusion.

However, when multiple people write modular sections of a paper, and some of those sections get highly technical, the result can be dull and disjointed prose. Additionally, I've had to continually remind myself while working on essay-reports not to miss the forest for the trees. My use of this idiom may seem odd. Distant reading strategies, including the scaled entity search used by the Arclight application, are supposed to help us see the larger pattern, to finally see the forest. But a focus on technical process and reporting results can also lead us away from the cultures, industries, and people that made us excited to study media history in the first place.

In her 2015 keynote talk at the Consortium of Humanities Centers and Institutes, historian Jill Lepore addressed these questions about writing quality and what we choose to study. "What humanists do is take the data we know about and put it into a language

that has force and beauty," Lepore said. "You can tell a story with numbers but using big data sometimes obscures the stories of those who are behind the numbers."[14] I believe that big data and Arclight's scaled entity search process can help contextualize and situate the stories that Lepore describes, but the risk she identifies about the data obscuring human agency certainly exists. Just as importantly, we should recognize that writing in a "language that has force and beauty" is essential to our work in the humanities. Forceful and elegant writing is not the only thing that matters—as this essay's first two sections hopefully make clear—but it is important nonetheless and something we need to preserve as we embark on expanded forms of scholarly production.

In my new book project, I am attempting to translate my data-intensive research out of the essay-report genre and into a ten-chapter book with forceful, compelling, and elegant prose. It has not been easy. The book, *Motion Papers: The Triumph of American Cinema's Trade Press*, explores the history of magazines that I have helped to digitize and make searchable. In particular, I am interested in why the film industry had more trade papers covering it than nearly any other American industry (over a dozen for most of the period from 1915 through 1940) and how each of the different publications operated within the industry. It is a big story; one that works best as a book—in which you can synthesize a great deal of information and allow an argument to slowly build—rather than as an essay, journal article, or web project.

My research has incorporated many of the same techniques of archival research and close reading that I used in my previous book, *Hollywood Vault: Film Libraries before Home Video*. I have found gems in the archival collections of the Margaret Herrick Library and Georgetown University Library. I have also spent a lot of time poking around court archives, which contain documents and testimonies from companies that never donated their papers to any institutional archive. Any lawsuit inherently has conflict,

and those conflicts can serve as engines for presenting history in dramatic, compelling ways. Beyond the archives, I spend quite a bit of time on my computer closely reading, keyword searching, and browsing through *Variety*, *Motion Picture News*, *Moving Picture World*, and the other trade papers. These familiar research processes leave me with impressions and theories. They also leave me with quotable lines of text that I can hold up as evidence to support my theories.

In *Motion Papers*, I have been trying to complement these research techniques with data analytics techniques, such as scaled entity search, topic modeling, and quantitative content analysis. These techniques have been tremendously helpful, even transformative, in observing changes that my close reading of the trade papers and archival documents could never reveal. Previous histories of *Variety*, for example, claimed that it increased its film industry reporting in the 1920s in response to film companies buying more advertising.[15] However, my quantitative analysis with Derek Long, Tony Tran, and Kit Hughes found that the inverse was true. It was only after devoting substantial resources to covering the film industry that *Variety* reaped the benefits of increases in film advertising.[16] This finding has informed my entire approach in *Motion Papers*. I understand the trade papers as active participants within a dynamic environment, making decisions about what to cover based on a range of strategies, journalistic assumptions, and industry changes.

When I carry out these quantitative methods, however, they do not leave me with quotable lines of text. Instead, they leave me with tables, graphs, data visualizations, abstractions. Alan Liu has observed that "one noticeable effect of distant reading in Moretti and Jockers's mode is that data visualizations of large patterns increasingly replace block quotations as the objects of sustained focus."[17] In my experience, I have found that this shift creates two major writing challenges for a book. First, it halts the momentum

of your argument and narrative because you need to stop to explain the methods used in the visualization. To skip this step is to make one's process opaque and create the proverbial "black box," which other scholars struggle to understand and come to regard with suspicion.[18] Second, even if you succeed at clearly and succinctly communicating your process, you still leave readers with an abstraction rather than a direct connection to someone who lived in the past and who wrote those quotable lines of text. This can have the effect Jill Lepore identified: "obscur[ing] the stories of those who are behind the numbers."[19] You may leave readers with more precise answers to certain quantifiable questions, but, on a broader level, a far less engaging and satisfying experience.

I have by no means resolved these conundrums. As I work on the book, the most effective strategy seems to be keeping the number of graphs to a minimum and only using them when they help propel the argument forward. Most of my graphs and visualizations will remain on my personal computer, just as most of the notes and digital photos I take at archives never make it into the book. The data visualizations, like the time spent immersing myself in an archival collection, helped me better understand my subject. But to present all of the visualizations in the book would be overwhelming and distract from the major points rather than clarifying them.

CONCLUSION

This essay has stressed the differences in the processes of curating, coding, and writing. However, there are many beneficial convergences too. The transitions can feel jarring as I move from packing up boxes of magazines to coding and executing Python scripts to writing intelligible sentences. But this mix can also be stimulating, even fun, and lead to interesting places. Tara McPherson has identified these benefits, advocating for multimodal scholars who "construct knowledge in and through our objects of study."[20] I feel fortunate to experience many of these moments.

For example, when I engage in the digitization and analysis of a year's worth of *Motion Picture Herald*—from a stack of magazines in my basement to digital scans to indexed metadata to data analytics and visualizations—I come away with deeper understandings of this particular publication and the decisions and transformations that are part of making a historic publication searchable online.

These insights and investigations only matter, though, to the extent that I can share them with others. One of the useful things about focusing on the divergences between curating, coding, and writing is that they lead us to the different expressions that these activities take: digital collections, software, publications. As humanities scholars, we sometimes resent it when a funding agency or administrator asks, "What's the deliverable?" It can seem too output oriented. Knowledge is never finished, always in progress. But we should also recognize that knowledge only continues forward when we express it in some form. As media historians, our work depends upon deliverables—the movies that producer Val Lewton delivered to RKO, for example, or the scripts and notes that Irna Phillips delivered to NBC. Producing the best work possible can take time. But I try to make sure I spend that time moving closer to a final deliverable that will speak to some audience, big or small.

If we take seriously digital collections, software, and publications as scholarly forms, we may come to witness another convergence—that between "research" and "service." These two categories typically have different places on a professor's CV, and the fact that service appears below both research and teaching is a reflection of its comparative status. I understand that serving on a departmental committee is not the same thing as carrying out a research program. But aren't the best research publications so impactful because they offer a service to the discipline and society at large? Similarly, if a digital collection or software tool presents

artifacts in a way that alters and advances our understanding of history, then should we not think of it as both a service and an expression of research? In the case of my own work, I suspect that none of the books and articles I write will match the transformative power of taking millions of pages of historic media publications—some canonical, most unknown—and putting them online for broad access along with tools and lenses to explore them. This may have something to do with my shortcomings as a writer. But I think it has much more to do with the tremendous opportunity we have right now to build openly accessible digital collections for the future, and to develop the software and methods that allow us to explore those collections in new ways.

ENDNOTES

1 See, for example, "Eric Hoyt - Data Mining Silent Cinema," YouTube video, 22:13, posted by "Eric Hoyt," July 23, 2014, https://www.youtube.com/watch?v=qO8W_ccIX7Y and Eric Hoyt, "How to Topic Model a Fan Magazine," *Project Arclight Blog*, last modified November 17, 2014, accessed February 12, 2016, http://projectarclight.org/arguments/how-to-topic-model-a-fan-magazine.

2 David Pierce and James Layton, *The Dawn of Technicolor: 1915–1935* (Rochester: George Eastman House Press, 2015); David Pierce, *The Survival of American Silent Feature Films: 1912–1929*, (Washington, DC: Council on Library and Information Resources and Library of Congress, 2013); David Pierce. "Forgotten Faces: Why Some of Our Cinema Heritage is Part of the Public Domain," *Film History* 19, no. 2 (2007): 125–43.

3 For more on the MHDL's background, see David Pierce, "Media History Digital Library," *Journal of Film Preservation* 88 (April 2013): 34–41; Eric Hoyt, Carl Hagenmaier, and Wendy Hagenmaier, "Media + History + Digital + Library: An Experiment in Synthesis," *Journal of Electronic Media Studies* 3, no. 1 (Spring 2013): accessed February 12, 2016, doi:10.1349/PS1.1938-6060.A.430.

4 "A Million Pictures: Magic Lantern Slide Heritage as Artefacts in the Common European History of Learning," accessed February, 12, 2016, http://a-million-pictures.wp.hum.uu.nl/.

5 Jerome McGann, “Philology in a New Key,” *Critical Inquiry* 39:2 (Winter 2013): 327.

6 Jerome McGann, “Philology in a New Key,” 338, 344.

7 For more on this, see Eric Hoyt, “Bootstrapping a Digital Archive? 5 Things to Consider,” *Spectator* 33, no. 2 (Fall 2013), 31–37.

8 Alan Galey, Stan Ruecker, and the INKE Research Group, “How a Prototype Argues,” *Literary and Linguistic Computing* 25, no. 4 (2010): 405–24. For more on the potential of software and databases to inform and convey arguments, see Tara McPherson, “Introduction: Media Studies and the Digital Humanities,” *Cinema Journal* 48, no. 2 (Winter 2009): 121–22; Stephen Ramsay and Geoffrey Rockwell, “Developing Things: Notes toward an Epistemology of Building in the Digital Humanities,” in *Debates in the Digital Humanities*, ed. Matthew K. Gold (Minneapolis: University of Minnesota Press, 2012), 75–84.

9 Lori Emerson, *Reading Writing Interfaces: From the Digital to the Bookbound* (Minneapolis: University of Minnesota Press, 2014), xviii.

10 I discuss this point further as it relates to Lantern in Eric Hoyt, “Lenses for Lantern: Data Mining, Visualization, and Excavating Film History’s Neglected Sources,” *Film History* 26, no. 2 (Summer 2014): 149–53.

11 Anne Friedberg, “On Digital Scholarship,” *Cinema Journal* 48, no. 2 (Winter 2009): 153.

12 Willard McCarty, “How to write an ‘essay-report’ in digital humanities,” last modified November 7, 2015, accessed February 12, 2016, https://docs.google.com/document/d/1mddTnCZ_w8JPKv-7OJ-iHa6r9wyuwzfhP-GJ1FXP9bIs/edit?pli=1.

13 Derek Long et al., “Who’s Trending in 1910s American Cinema? Exploring ECHO and MHDL at Scale with Arclight,” *The Moving Image* (forthcoming); Eric Hoyt et al., “*Variety*’s Transformations: Digitizing and Analyzing the First Forty Years of the Canonical Trade Paper,” *Film History* 27, no. 4 (2015): 75–106; Kit Hughes et al., “Hacking Radio History’s Data: Station Call Letter, Digitized Magazines, and Scaled Entity Search,” *Media Industries Journal* 2, no. 2 (2015), http://www.mediaindustriesjournal.org/index.php/mij/article/view/128/182; Eric Hoyt et al., “Scaled Entity Search: A Method for Media Historiography and Response to Critiques of Big Humanities Data Research,” *Proceedings of IEEE Conference on*

Big Data (2014): 51–59, http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7004453; Eric Hoyt, Kevin Ponto, and Carrie Roy, "Visualizing and Analyzing the Hollywood Screenplay with ScripThreads," *Digital Humanities Quarterly* 8, no. 4 (2014), http://www.digitalhumanities.org/dhq/vol/8/4/000190/000190.html.

14 Jill Lepore quoted in Mary Ellen Gabriel, "Humanities by the Numbers: Global Consortium Meets at UW-Madison," *University of Wisconsin–Madison College of Letters and Sciences News*, last modified June 15, 2015, accessed February 12, 2016, http://news.ls.wisc.edu/humanities-the-arts/humanities-by-the-numbersglobal-consortium-meets-at-uw-madison/.

15 Peter Besas, *Inside* Variety*: The Story of the Bible of Show Business, 1905–1987* (Madrid: Ars Millenii, 2000), 187; Dayton Stoddart, *Lord Broadway:* Variety*'s Sime* (New York: Wilfred Funk, 1941), 129.

16 Eric Hoyt et al., "*Variety*'s Transformations," 96.

17 Alan Liu, "Where Is Cultural Criticism in the Digital Humanities?" in *Debates in the Digital Humanities*, ed. Matthew K. Gold (Minneapolis: University of Minnesota Press, 2012), 494.

18 "Black box" ranks up with "distant reading" as one of the most used terms in the digital humanities. For some insightful works that discuss the black box, see Matthew G. Kirschenbaum, *Mechanisms: New Media and the Forensic Imagination* (Cambridge: MIT Press, 2008); Fred Gibbs and Trevor Owens, "Building Better Digital Humanities Tools: Toward Broader Audiences and User-Centered Designs," *Digital Humanities Quarterly* 6, no. 2 (2012): accessed February 12, 2016, http://www.digitalhumanities.org/dhq/vol/6/2/000136/000136.html.

19 Jill Lepore quoted in Mary Ellen Gabriel, "Humanities by the Numbers."

20 Tara McPherson, "Introduction," 120.

BIBLIOGRAPHY

Besas, Peter. *Inside "Variety": The Story of the Bible of Show Business, 1905–1987*. Madrid: Ars Millenii, 2000.

Emerson, Lori. *Reading Writing Interfaces: From the Digital to the Bookbound*. Minneapolis: University of Minnesota Press, 2014.

"Eric Hoyt - Data Mining Silent Cinema." YouTube video, 22:13. Post-

ed by “Eric Hoyt.” July 23, 2014. https://www.youtube.com/watch?v=qO8W_ccIX7Y.

Friedberg, Anne. “On Digital Scholarship.” *Cinema Journal* 48, no. 2 (Winter 2009): 150–54.

Gabriel, Mary Ellen. “Humanities by the Numbers: Global Consortium Meets at UW-Madison,” *University of Wisconsin–Madison College of Letters and Sciences News*. Last modified June 15, 2015. Last accessed February 12, 2016. http://news.ls.wisc.edu/humanities-the-arts/humanities-by-the-numbersglobal-consortium-meets-at-uw-madison/.

Galey, Alan, Stan Ruecker, and the INKE Research Group. “How a Prototype Argues.” *Literary and Linguistic Computing* 25, no. 4 (2010): 405–24.

Gibbs, Fred and Trevor Owens. “Building Better Digital Humanities Tools: Toward Broader Audiences and User-Centered Designs.” *Digital Humanities Quarterly* 6, no. 2 (2012). Accessed February 12, 2016. http://www.digitalhumanities.org/dhq/vol/6/2/000136/000136.html.

Hoyt, Eric. “Bootstrapping a Digital Archive? 5 Things to Consider.” *Spectator* 33, no. 2 (Fall 2013): 31–37

Hoyt, Eric. “How to Topic Model a Fan Magazine.” *Project Arclight Blog*. Last modified November 17, 2014. Accessed February 12, 2016. http://projectarclight.org/arguments/how-to-topic-model-a-fan-magazine.

Hoyt, Eric. “Lenses for Lantern: Data Mining, Visualization, and Excavating Film History’s Neglected Sources.” *Film History* 26, no. 2 (Summer 2014): 146–68.

Hoyt, Eric, Carl Hagenmaier, and Wendy Hagenmaier. “Media + History + Digital + Library: An Experiment in Synthesis.” *Journal of Electronic Media Studies* 3, no. 1 (Spring 2013). Accessed February 12, 2016. doi:10.1349/PS1.1938-6060.A.430.

Hoyt, Eric, Derek Long, Anthony Tran, and Kit Hughes. “*Variety*’s Transformations: Digitizing and Analyzing the First Forty Years of the Canonical Trade Paper.” *Film History* 27, no. 4 (2015): 75–106.

Hoyt, Eric, Kevin Ponto, and Carrie Roy. “Visualizing and Analyzing the Hollywood Screenplay with ScripThreads.” *Digital Humanities*

Quarterly 8, no. 4 (2014). http://www.digitalhumanities.org/dhq/vol/8/4/000190/000190.html.

Hoyt, Eric, Kit Hughes, Derek Long, Kevin Ponto, and Anthony Tran. "Scaled Entity Search: A Method for Media Historiography and Response to Critiques of Big Humanities Data Research." *Proceedings of IEEE Conference on Big Data* (2014): 51–59. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7004453.

Hughes, Kit, Eric Hoyt, Derek Long, Kevin Ponto, and Tony Tran. "Hacking Radio History's Data: Station Call Letter, Digitized Magazines, and Scaled Entity Search." *Media Industries Journal* 2, no. 2 (2015). http://www.mediaindustriesjournal.org/index.php/mij/article/view/128/182

Kirschenbaum, Matthew G. *Mechanisms: New Media and the Forensic Imagination*. Cambridge: MIT Press, 2008.

Liu, Alan. "Where Is Cultural Criticism in the Digital Humanities?" In *Debates in the Digital Humanities*, edited by Matthew K. Gold, 490–510. Minneapolis: University of Minnesota Press, 2012.

Long, Derek, Eric Hoyt, Anthony Tran, Kevin Ponto, and Kit Hughes. "Who's Trending in 1910s American Cinema? Exploring ECHO and MHDL at Scale with Arclight." *The Moving Image* (forthcoming).

McCarty, Willard. "How to write an 'essay-report' in digital humanities," Last modified November 21, 2015. Accessed February 12, 2016. https://docs.google.com/document/d/1mddTnCZ_w8JPKv-7OJ-iHa6r-9wyuwzfhPGJ1FXP9bIs/edit?pli=1.

McGann, Jerome. "Philology in a New Key." *Critical Inquiry* 39, no. 2 (Winter 2013): 327–46.

McPherson, Tara. "Introduction: Media Studies and the Digital Humanities." *Cinema Journal* 48, no. 2 (Winter 2009): 119–23.

Million Pictures, A. "A Million Pictures: Magic Lantern Slide Heritage as Artefacts in the Common European History of Learning." Accessed February, 12, 2016, http://a-million-pictures.wp.hum.uu.nl/.

Pierce, David. "Forgotten Faces: Why Some of Our Cinema Heritage is Part of the Public Domain." *Film History* 19, no. 2 (2007): 125–43.

Pierce, David. "Media History Digital Library." *Journal of Film Preservation* 88 (April 2013): 34–41.

Pierce, David. *The Survival of American Silent Feature Films: 1912–1929*. Washington, DC: Council on Library and Information Resources and Library of Congress, 2013.

Pierce, David and James Layton. *The Dawn of Technicolor: 1915–1935*. Rochester: George Eastman House Press, 2015.

Ramsay, Stephen and Geoffrey Rockwell. "Developing Things: Notes toward an Epistemology of Building in the Digital Humanities." In *Debates in the Digital Humanities*, edited by Matthew K. Gold, 75–84. Minneapolis: University of Minnesota Press, 2012.

Stoddart, Dayton. *Lord Broadway:* Variety*'s Sime*. New York: Wilfred Funk, 1941.

KEYWORDS AND ONLINE RESOURCES

Robert Hunt and Tony Tran

ACTION Toolkit
ACTION (Audio-visual Cinematic Toolbox for Interaction, Organization, and Navigation) is an open source platform that supports the computational analysis of film and other audiovisual materials. Elements of analysis include audio and color features, motion, and structural segmentations. http://digitalhumanities.dartmouth.edu/projects/the-action-toolbox/

Algorithm
A series of steps that a human or computer can follow in order to solve a problem or carry out a process.

AntConc
A software program that allows users to create concordances and conduct textual analysis. http://www.laurenceanthony.net/software/antconc/

ArcGIS
A platform used to create, manage, and present geographical data with maps. https://www.arcgis.com/

Arclight
Project Arclight is a data mining and visualization tool for film and media history that allows users to analyze millions of pages of digitally scanned newspapers and magazines. http://projectarclight.org/

Black Box
In digital humanities, a black box refers to the underlying software or algorithms of platforms that are not fully understood or seen by users, either through a lack of technical knowledge or the inability to inspect the platform.

Boolean Operators
Boolean Operators are words used to combine or exclude keywords in searches. Examples include AND, OR, NOT, and AND NOT.

Britain on Film
A digital archive and resource focused on British lives and film.
http://www.bfi.org.uk/britain-on-film

CESIF
The Canadian Educational, Sponsored, and Industrial Film (CESIF) Project is an online database of information about Canadian film titles, originating from private production outfits, in the broadly defined genres of educational, sponsored, and industrial motion pictures.
http://www.screenculture.org/cesif/

Chronicling America: Historic American Newspapers
The Library of Congress's Chronicling America website provides a database of information about US newspapers as well as select digitized newspaper pages. http://chroniclingamerica.loc.gov/

Cinema Tools Program
A software program designed for filmmakers to use with the editing program Final Cut Pro that allows users to create databases of film material.

Cinema Treasures
A crowd-sourced database of historical and contemporary movie theaters that can be plotted on a Google map by country.
http://cinematreasures.org/

Cinematographic Atlas of Canadian Movie Theatres
Similar to Cinema Treasures, this website allows users to plot and analyze Canadian theaters on maps.
http://atlascine3.classone-tech.com/index.html

Cinemetrics
A movie measurement and study tool and database that collects statistics on audiovisual materials, including average shot length, number of shots, and types of shots. http://www.cinemetrics.lv/

Concordance
A list of words in a text or set of texts that also typically shows the context around these words.

Corpus
A collection or body of data. A corpus can be constructed by the researcher or defined by an archive, a library, a database, and/or the availability of materials.

CSS
Cascading Style Sheets are used to control the visual layout of HTML code on a website.

CSV
Comma Separated Values files store data (text and numbers) in a format where each entry or field is separated by a comma. CSV files are frequently opened and edited using Excel or other spreadsheet software programs.

Data
Data can be defined in a number of ways. In a digital humanities context, data can be understood as information that can be systematically collected, organized, and analyzed, often with the aid of a computer.

Data Visualization
Data visualization is the process of presenting data or information in a visual context, such as a graph or chart, to make it easier to understand.

Database
A collection of information organized so that it can be easily accessed (for

example, by a computer program).

Dataset
An individual collection of organized data or information.

dBase
One of the first database management systems for computers.

Denmark on Film
A web portal for the Danish Film Institute that covers 1905–65 and includes an interactive map with films from specific regions and cities.
http://filmcentralen.dk/museum/danmark-paa-film/kort

DEVONthink
An information and document management and retrieval program, DEVONthink also has the ability to connect and organize documents based on relevant topic words.
http://www.devontechnologies.com/products/DEVONthink/overview.html

Distant Reading
A term originating from the work of literature scholar Franco Moretti, "distant reading" is the process of aggregating and analyzing large amounts of data to see trends or patterns across several datasets. This is often contrasted with "close reading," a focus on individual or relatively few objects for textual analysis.

ECHO
Early Cinema History Online is a filmographic database featuring credits for over 35,000 titles released in the US from 1908 to 1920.
http://echo.commarts.wisc.edu

Entity List
A list of terms or items of interest that is complied by a researcher and processed through an application or software.

Field Guide to Sponsored Films
Rick Prelinger's freely available guide to hundreds of notable American sponsored films, created in 2006.
http://www.filmpreservation.org/userfiles/image/PDFs/sponsored.pdf

Final Cut Pro
A nonlinear video editing software program developed by Macromedia, Inc. and Apple, Inc.

Free Picture Resizer
A tool that can batch convert and resize digital photographs and images.
http://download.cnet.com/Free-Picture-Resizer/3000-12511_4-10297789.html

Free Studio 5
A freeware multimedia package toolkit that can convert audio and video files to different formats.
http://www.dvdvideosoft.com/free-dvd-video-software.htm

GIS
A Geographic Information System is a system used to record, manage, analyze, and visually present different types of spatial and geographical data.

Going to the Show
Going to the Show documents the experience of moviegoing in North Carolina from 1896 to the 1930s. Employing maps, newspaper ads, photographs, city directories, and more, Going to the Show explores the intersections of moviegoing with race and urban and rural life.
http://docsouth.unc.edu/gtts/map/

GPS
Global Positioning System is a satellite-based navigation system that provides location information.

Handbrake
A free and open source video transcoder that allows users to convert video to different formats.

HathiTrust Digital Library
The product of a partnership between numerous academic and research institutions, this online collection offers millions of titles digitized from libraries around the world. https://www.hathitrust.org/

Heat Maps
A graphical representation of data where individual values in a grid are represented as colors, often used to visualize frequency or intensity.

HoMER
The History of Movie-going, Exhibition, and Reception Project is a collective that promotes research on the international phenomena of film exhibition and reception, with an emphasis on searchable databases, graphic imaging software, and digital formats. http://homernetwork.org/

HTML
Hyper Text Markup Language is a markup language used to create web pages.

ImageJ
A Java image processing program. Some functions include measuring distances and angles, image manipulations, and geometric transformations.

[In]Transition
A collaboration between MediaCommons and the Society for Cinema and Media Studies' *Cinema Journal, [in]Transition* is a peer-reviewed academic journal of video essays that present film and moving image studies research. http://mediacommons.futureofthebook.org/intransition/

IPS
Integrated Preservation Software is a database program created by Larry

Karr for the National Park Service. The database contains surveys and information on approximately 380 properties in the Washington Heights area of the District of Columbia.

iSkysoft iTube Studio
A tool to batch download YouTube and other online videos and convert them to different formats. https://www.iskysoft.com/

JavaScript
A programing language often used to program behaviors and actions of web pages.

JotNot
An iPhone and iPad document scanner application that uses the device's camera to capture documents. http://www.jotnot.com/

jQuery
jQuery is a JavaScript library designed to simplify web page functions and coding by using the user's browser.

Kinomatics Project
The Kinomatics Project collects, analyzes, and visualizes data about the creative industries, including the locations and times of international film exhibition. http://kinomatics.com/

Lantern
A search and visualization platform for the collections of the Media History Digital Library, which includes nearly 2 million pages of digitized books and periodicals related to the histories of cinema, broadcasting, and recorded sound. http://lantern.mediahist.org

MALLET
A software program that allows a user to perform topic modeling. http://mallet.cs.umass.edu/topics.php

Mapping Movies
Mapping Movies allows users to digitally explore changing landscapes of social and spatial history by investigating the locations and movements of moving pictures. http://mappingmovies.unh.edu/maps/erma.html

Mapping the City in Film
The Liverpool - City in Film map features location data relating to 176 former cinema sites across Merseyside. It also features a selection of embedded videos of digitized films of the city made by amateur filmmakers and others dating back to the 1930s.
https://www.liv.ac.uk/architecture/research/cava/cityfilm/map/

Markov Chain
Mathematical systems that jump from one state, or set of values, to another state.

MediaCommons
An online community of scholars and practitioners of media studies who explore new ways to publish work within the field.
http://mediacommons.futureofthebook.org/

Mediathread
A communal online media collection of source materials and assignments for students and instructors that allows users to annotate, organize, and share media. http://mediathread.columbia.edu/

MEP
The Media Ecology Project is a digital resource at Dartmouth that enables researchers to digitally access archival moving image collections and contribute back to the archival and research communities through the fluid contribution of metadata and other knowledge.
https://sites.dartmouth.edu/mediaecology/

Metadata
Metadata is data that describes other data. It can provide basic descrip-

tions of the content, structure, and context of data (e.g., date of creation, type of file, size of file, etc.). Tagging is a form of metadata.

MHDL
The Media History Digital Library digitizes magazines and periodicals related to the histories of cinema, broadcasting, and sound. http://mediahistoryproject.org/

MIRC
The University of South Carolina's Moving Image Research Collection is a wide-ranging digital repository for rare archival film and media materials from all over the globe, including Chinese films, US regional home movie collections, and science and nature films. http://mirc.sc.edu/

MP4
MPEG-4 is a digital multimedia format used to store audio and video.

MySQL
A popular open source database often used for web applications. https://www.mysql.com/

Neatline
Neatline allows users to create stories with interactive digital maps and timelines. http://neatline.org/

Ngram Viewer
Google's Ngram Viewer is a search engine that charts the frequency of entities in Google's corpora of scanned texts between 1500 and 2008. https://books.google.com/ngrams

OCR
Optical Character Recognition is a process of converting printed text into digital or machine-readable text.

Omeka
A free and open source web-publishing platform designed for archives, museums, and scholarly collections and exhibitions. http://omeka.org/

Onomy
A website where you can create and share taxonomies, folksonomies, and other forms of controlled vocabularies for use on the semantic web. http://onomy.org

Open Archives Initiative
The Open Archives Initiative promotes interoperability standards that encourage the efficient dissemination of content. https://www.openarchives.org/

ORBIS
ORBIS: The Stanford Geospatial Network Model of the Roman World is an interactive map of the Roman world circa 200 CE that allows users to explore transportation networks and routes. http://orbis.stanford.edu/

Perl
A programming language that is especially useful for the manipulation of text files.

Photoshop
An Adobe software product used primarily to edit images.

PHP
PHP is a server scripting language often used to make dynamic and interactive web pages.

Project Bamboo
Project Bamboo was a cyberinfrastructure initiative for arts and humanities research that aimed to develop shared technology services; it ended in 2012. http://www.projectbamboo.org/

Python
A programming language used for web programming, software development, and interface development.

QuickTime Player 7
A media player software application created by Apple, Inc.

Red Hen Lab
A cooperative of researchers working on developing theories and digital tools to assist research into multimodal communication.
https://sites.google.com/site/distributedlittleredhen/home

Ripping
The process of copying audio-visual content from one form (DVD, online, etc.) to another platform.

Ruby on Rails
A web application development framework written in the Ruby programming language. It is designed to make web coding more efficient.
http://rubyonrails.org/

Scalar
A free and open source publishing platform that emphasizes long-form digital scholarship. Scalar allows users to assemble media and text to produce arguments and experiments with online academic work.
http://scalar.usc.edu/scalar/

Scale
The range of values or data included in a corpus or study. Distant reading and big data often employ large scales of study by looking at massive amounts of data, while close readings have smaller scales by focusing on specific objects of study.

Scapple
A free-form text editor that allows users to make notes and link them

with lines and arrows. https://www.literatureandlatte.com/scapple.php

Scripts
Computer scripts are the software code that run and execute programs.

Scrivener
A word processor and project management tool that allows users to take notes, view research alongside written text, and edit a document. https://www.literatureandlatte.com/scrivener.php

SES
Scaled Entity Search is a method that allows users to search for hundreds or thousands of queries and analyze their trends across a corpus simultaneously. SES also emphasizes an analytical framework that considers the relationships among the entities, the corpus, and digital technologies to interpret the results fully.

Software Studies Initiative
A research lab and a design studio working on the analysis of big cultural datasets, resulting in data visualizations, interactive installations, free software tools, and research papers and books. http://lab.softwarestudies.com/

Solr
An open source search and indexing platform that powers search and navigation functions for websites. http://lucene.apache.org/solr/

SQL
Structured Query Language is a programing language designed for managing data in relational databases.

Stack
Within ImageJ, a stack is a set or collection of related images within one window.

Stop Words/List
A list of commonly used words to ignore when performing text mining.

Story Maps
Story Maps employs maps and geo-tagging to let users tell sequential and place-based narratives. https://storymaps.arcgis.com/en/

String Literals
String literals are data or characters enclosed in single or double quotes. They are often used to enclose text and helpful when texts contain punctuation.

Tagging
The assigning of a keyword or term to another piece of data. Tags are a form of metadata.

Text Mining
A digital form of text analytics where digital text is processed and analyzed to look for patterns, trends, and statistical data. This may include exploring word frequencies, word placement, and sentence structures.

TextDNA
A tool that lets users analyze and compare word usage across text collections of varying scales.
http://graphics.cs.wisc.edu/Vis/SequenceSurveyor/TextDNA.html

TextSTAT
Text Simple Text Analysis Tool (STAT) is a program for analyzing texts. It produces word frequency lists and concordances.
http://neon.niederlandistik.fu-berlin.de/en/textstat/

Topic Modeling
A process, usually performed with the aid of specific topic-modeling digital tools, whereby a corpus of texts is analyzed to reveal its likely topics. Typically, the computer looks for clusters of words that frequently reoc-

cur. It then makes a probabilistic assessment about which word clusters constitute a topic and about their relative prominence.

Tumblr
A microblogging and social media platform. http://www.tumblr.com

UNIX
Developed in the 1960s, UNIX is a computer operating system, or a suite of programs that make computers operate.

VEMI Lab
The Virtual Environment and Multimodal Interaction Laboratory is part of the Spatial Informatics program in the School of Computing and Information Science at the University of Maine. The program's mission is to study and design technologies aimed towards blind/visually impaired people. http://www.vemilab.org/

Video Analysis Tableau
An online toolkit designed for automated video comparison, annotation, and visualization. http://thevatproject.org/

VisualHub
A video converter that makes audiovisual media compatible with different devices.

Voyant
A web-based textual analysis tool. http://voyant-tools.org/

W3C Open Annotation Format
A movement to develop a set of specifications and standards for an interoperable Web annotation system and architecture. This includes the ability to annotate and highlight web pages, e-books, videos, audio streams, and more.

Wiki
A website that allows users to create and modify its content collaboratively.

Women Film Pioneers Project
The WFPP is an online database that highlights the hundreds of women who worked behind the scenes in the silent film industry as directors, producers, editors, and more. The database features career profiles, essays, still and moving images, and archival and bibliographic resource materials. https://wfpp.cdrs.columbia.edu/

WordPress
A website for users to create a website or blog with relative ease. https://wordpress.com/

XML
EXtensible Markup Language is designed to store, transport, and exchange data with a focus on describing the data.

xQuery
A computing language used to search XML data.

Z-Projections
Within ImageJ, z-projections are a method of manipulating images by placing them on top of each other (on the *z* axis).

Zotero
A reference management software program used to create and organize bibliographic data. https://www.zotero.org/

CONTRIBUTORS

Charles R. Acland is professor of communication studies at Concordia University, Montreal, and codirector of the Media History Research Centre. He is the author of *Swift Viewing: The Popular Life of Subliminal Influence* and coeditor, with Haidee Wasson, of *Useful Cinema*. He is codirector of Project Arclight, with Eric Hoyt.

Kevin L. Ferguson is assistant professor at Queens College, City University of New York, where he teaches teaches college writing, contemporary literature, digital humanities, and film adaptation. His book *Eighties People* examines new cultural figures in the American 1980s.

Kaitlin Fyfe is a graduate student in film studies at the University of Wisconsin–Madison where she teaches film editing.

Laura Horak is assistant professor of film studies at Carleton University. She is the author of *Girls Will Be Boys: Cross-Dressed Women, Lesbians, and American Cinema, 1908–1934* and coeditor of *Silent Cinema and the Politics of Space*, winner of the 2015 Society for Cinema and Media Studies Award for Best Edited Collection.

Eric Hoyt is assistant professor of communication arts at the University of Wisconsin–Madison. He is the author of *Hollywood Vault: Film Libraries before Home Video*, codirector of the Media History Digital Library (http://mediahistoryproject.org), lead developer of Lantern (http://lantern.mediahist.org), and codirector of Project Arclight (http://projectarclight.org). His current book project combines archival research, close reading, and computational distant reading to examine the history of the American film industry's trade press.

Kit Hughes is assistant professor of communication studies at Colorado State University, where she researches industrial media, television history, and the politics of archives. Her book-length project examines the rise of industrial television in postwar America. Her work has appeared in *Media, Culture & Society*; *Television & New Media*; *American Archivist*; and *Film Criticism*.

Robert Hunt is a graduate student in the Department of Communication Studies at Concordia University, Montreal.

Lea Jacobs teaches film history and aesthetics at the University of Wisconsin–Madison and is the author of *The Wages of Sin: Censorship and the Fallen Woman Film*, *Theatre to Cinema* (cowritten with Ben Brewster), *The Decline of Sentiment: American Film in the 1920s*, and *Film Rhythm After Sound: Technology, Music and Performance*.

Elana Levine is associate professor in the Department of Journalism, Advertising, and Media Studies at the University of Wisconsin–Milwaukee. She is the author of *Wallowing in Sex: The New Sexual Culture of 1970s American Television*, coauthor of *Legitimating Television: Media Convergence and Cultural Status*, coeditor of *Undead TV: Essays on "Buffy the Vampire Slayer,"* and the editor of *Cupcakes, Pinterest, and Ladyporn: Feminized Popular Culture in the Early Twenty-first Century*.

Derek Long is a PhD candidate in film at the University of Wisconsin–Madison. He is the data manager, curator, and analyst for Project Arclight's US team, and he is the project manager for Early Cinema History Online (ECHO). His dissertation examines distribution and production planning in the early years of the Hollywood studio system, and his work has been published in the *Historical Journal of Film, Radio, and Television* and the *New Review of Film and Television Studies*.

Fenwick McKelvey is assistant professor of communication studies at Concordia University, Montreal. His work has appeared in numerous journals, including *Television and New Media*, the *International Journal of Communication*, the *European Journal of Cultural Studies*, the *Canadian Journal of Communication*, *Global Media Journal*, and the *Journal of Information Technology and Politics*. He is coauthor of *The Permanent Campaign: New Media, New Politics*. His book *Media Daemons* is currently under contract with the University of Minnesota Press.

Cynthia B. Meyers is the author of *A Word from Our Sponsor: Admen, Advertising, and the Golden Age of Radio* and associate professor of communication at the College of Mount Saint Vincent, Bronx, New York. Her articles have appeared in the *Historical Journal of Film, Radio and Television*, the *Journal of Radio and Audio Media*, *Cinema Journal*, and book collections such as *Media Industries: History, Theory, and Method* and *Films That Sell*.

Miriam Posner is digital humanities program coordinator and a member of the core digital humanities faculty at the University of California, Los Angeles. Her book *Depth Perception*, on medical filmmaking, is under contract with the University of North Carolina Press. She is also a member of the executive council of the Association for Computers and the Humanities.

Lisa Spiro is executive director of digital scholarship services at Rice University's Fondren Library. She has contributed chapters to *Digital Humanities Pedagogy*, *Debates in the Digital Humanities*, *Collaborative Approaches to the Digital in English Studies*, and *#alt-academy: Alternate Academic Careers for Humanities Scholars*. She served as chair of the communications committee for the Alliance of Digital Humanities Organizations and is currently a member of the steering committee for the Texas Digital Humanities Consortium.

Tony Tran is a PhD candidate in media and cultural studies in the Department of Communication Arts at the University of Wisconsin–Madison. His research focuses on everyday digital life in Vietnamese diasporic urban communities.

Deb Verhoeven is chair and professor of media and communication at Deakin University and the director of the Humanities Networked Infrastructure (HuNI) project (http://huni.net.au). She served as inaugural deputy chair of the National Film and Sound Archive of Australia from 2008 to 2011 and as CEO of the Australian Film Institute from 2000 to 2002. She is the director of the Kinomatics Project (http://kinomatics.com), an interdisciplinary study that collects, explores, analyzes, and represents data about the creative industries.

Gregory A. Waller is professor and chair of cinema and media studies in the Media School at Indiana University. He is currently the editor of *Film History*. His publications include *Main Street Amusements: Film and Commercial Entertainment in a Southern City, 1895–1930*, and his recent research focuses on the history of nontheatrical cinema.

Haidee Wasson is professor of film studies at Concordia University, Montreal. Her books include *Museum Movies: The Museum of Modern Art and the Birth of Art Cinema*, *Inventing Film Studies* (coedited with Lee Grieveson), and *Useful Cinema* (coedited with Charles R. Acland). She is coeditor of the Cultural Histories of Cinema Series for the British Film Institute. Her current research investigates the history of portable film projectors.

Mark Williams is associate professor of film and media studies at Dartmouth College and director of the Media Ecology Project. He has published in a variety of journals and anthologies, including *Télévision: le moment expérimental (1935–1955)*; *Convergence Media History*; *Collecting Visible Evidence*; *Dietrich Icon*; and *Living*

Color: Race, Feminism, and Television. He is the founding editor of the e-journal *Journal of e-Media Studies* and coeditor of the book series Interfaces: Studies in Visual Culture for the University Press of New England.

INDEX

Acland, Charles R. 7, 21, 22, 24, 238, 351
ACTION Toolkit *see* Audio-visual Cinematic Toolbox for Interaction, Organization, and Navigation
Adobe
    Acrobat 306–307
    Creative Cloud 347
    Photoshop 250, 253, 274–275, 347, 383
    Premiere ProCC 250, 347
Advertising 2, 23, 46–47, 50, 59, 240, 309, 318, 320–321, 324, 326, 365
Algorithm 3–4, 14, 19, 21, 23–24, 31, 38, 40, 133, 135, 142, 145, 154, 172, 189, 279, 291, 359, 374
Allen, Robert C. 69, 302, 352
Alliance for Networked Visual Culture 337
Alliance of Digital Humanities Organizations 2, 391
annotation 253, 341, 342, 343, 387 *see* also W3C Open Annotation
AntConc 218–220, 227, 374
Apple, Inc. 378, 384
ArcGIS 66, 76, 82, 85–87, 374, 386
Arclight 4, 15–20, 22, 45, 103–11, 115–16, 118, 120, 121, 123, 125, 239–40, 243–247, 358–59, 363–64, 374
    Guidebook 1, 5, 7, 9, 10, 19, 24
    Project 11–13, 19, 150, 238, 363
    Symposium 197
Arrowsmith, Colin 68, 73, 81
ART+COM Studios 279
Audio-visual Cinematic Toolbox for Interaction, Organization, and Navigation (ACTION Toolkit) 374
Barker, Reginald 149–151
big data *see* data
big dataset *see* dataset
Black Box/black box 14, 16, 228, 366, 374
blockbuster 238–239
Boolean Operators 152, 375

box office 22, 114, 238, 240–42, 245–46
Box Office Attractions Company 157
Britain on Film 71, 375
broadcasting 2, 147, 196, 318, 320, 325, 338–39, 344, 347, 349, 354, 380, 382
Canadian cinema *see* cinema
Canadian Educational, Sponsored, and Industrial Film Project (CESIF) 21, 36, 137–40, 351, 375
Caquard, Sebastién 67–68, 71–72, 78
Carleton University 66–87
Cascading Style Sheets (CSS) 85, 187, 358, 376
Center for Spatial Studies 74
CESIF *see* Canadian Educational, Sponsored, and Industrial Film Project
Chicago 50, 103, 118–19, 122, 355
Chinese films 354, 382
Cinema
    data 166, 169, 175
    history 12, 34–35, 79, 83, 146, 270
    multi-sited 32, 46–47, 57, 60
    national 139, 296, 339, 356
    nontheatrical 45–50, 57, 59–60, 106, 392
Cinema Tools Program 251, 375
Cinema Treasures 375
Cinematic
    apparatus 32
    cartography/cartographies 67–68, 73
    representations 71–72
Cinematographic Atlas of Canadian Movie Theatres 70, 375
Cinemetrics 7, 249, 268, 280, 376
close mining 137
close reading 103, 110, 200, 202, 217, 342, 344, 364–65, 377, 384
coding 23, 158, 197, 307, 342, 358, 366, 367, 380, 384
Columbia University 337, 356
Comma Separated Value (CSV) 15, 110, 154, 253, 376
concordance 22, 218–220, 224, 374, 376, 386
Concordia University 11, 13, 137

controlled vocabularies 185, 340, 383
corpus 11, 16–19, 103–04, 112, 114–18, 121, 124–25, 135–37, 139–40, 152, 161, 204, 207, 218–20, 223–27, 231, 238–40, 244, 247, 277, 281, 287, 289–91, 294–96, 376, 384–86
CSS *see* Cascading Style Sheets
CSV *see* Comma Separated Values
Cultural Atlas of Australia 73
curating 23, 341, 347–50, 367
cybercartography 66–67
Dartmouth College 336–38
Data
    analytics 365
    big 11, 13, 15, 33, 37, 45, 135, 141–42, 165–70, 172, 175, 177, 268, 364, 384
    big cultural 177, 385
    management system 306–07
    metadata 7, 15–16, 137, 145–54, 159–60, 206, 306–07, 209, 326, 336–338, 340–44, 349–50, 360, 367, 381–82, 386
    mining 374 *see also* close mining; text mining
    small 22, 60, 249, 268
    visualization 7, 45, 74, 86, 172, 238, 365–67, 376, 385
database 21, 36, 40, 71, 140, 167, 176–77, 184–90, 249, 301, 306–07, 309, 312, 339, 352, 354, 356, 358, 375–77, 379–80, 382, 385, 388
    management system 148, 377
dataset 1, 4, 21, 67, 69, 85, 87
dBase 148–51, 377
Deakin University 171
Denmark on Film 377
DEVONthink 22, 306–316, 377
Digital
    humanities 1–4, 6–7 10–15, 19
    geospatial 20–23, 65–67, 73, 78, 80, 82, 133, 141, 157, 184, 186, 191, 197, 271–72, 279–80, 343, 347–48, 36–61, 374, 376
    Digital Humanities Now 6
    Digital Scholarship Lab 79

map 2, 7, 72, 80–81
mapping 65–68, 78
multimedia 84, 382
print archives 45, 58
repository 176, 336, 354, 382
scholarship 10, 79, 140, 153, 249, 279, 339–40, 384
surrealism 27–74, 279–80, 291
technologies 1, 3, 5–7, 19, 66, 178, 385
Dirichlet allocation (LDA) 14, 224
discussion/message boards 199
distant reading 18, 103, 158, 201–02, 280, 344, 365, 377, 384
DOMITOR 338
Drucker, Johanna 11, 80, 186–87
Drupal 138
Early Cinema History Online (ECHO) 21, 146, 350, 377
Edmonton Pipelines 81
entity list 16, 104–08, 115–16, 377
Ferguson, Kevin L. 1, 7–8, 22–23
filmographic database 146, 352, 377
Final Cut Pro (FCP) 250, 263, 375, 378
Flow and Movement 75
Free Picture Resizer 201, 378
Free Studio 5 201, 378
Fyfe, Kaitlin 7–8, 22, 249
Gender Research Institute 338
Geographic Information System (GIS) 5, 65–69, 72, 76, 78, 83, 87, 105, 123, 378
Geomatics and Cybercartographic Research Center 82
Georgetown University 338, 364
Ghosts of the Horseshoe 5
GIS *see* Geographic Information System
Global Flow of People 77
Global Positioning System (GPS) 5, 66–67, 378
Going to the Show 69, 352, 378
Google

Analytics 327, 360
Books 16, 220
Earth 66, 83
Maps/map 66, 69–70, 74, 80, 86, 189, 375
GPS *see* Global Positioning System
Handbrake 303, 379
Harvard University 50, 79
HathiTrust Digital Library 1, 59, 379
historiography
television 302, 314
see also media
History of Moviegoing, Exhibition and Reception (HoMER) 69, 379
Hollywood 2, 13, 22, 106, 119, 122, 125, 154, 158–59, 218, 223, 229, 239, 243, 248, 356
Horak, Laura 7, 20–21, 23
Hoyt, Eric 4, 17, 23, 45, 217, 352, 354
HTML *see* Hyper Text Markup Language
Hughes, Kit 1, 16, 19–21, 343, 365
Humanities+Design Lab 75
Hyper Text Markup Language (HTML) 85, 187, 207, 358, 361, 376, 379
ImageJ 22, 196, 200–04, 277–78, 281, 292, 279, 285, 388
India 113, 339
instantiation 46, 50, 53, 57
Integrated Preservation Software (IPS) 379
[In]Transition 8, 9, 348, 379
iSkysoft iTube Studio 303, 380
Jacobs, Lea 7–8, 22, 249
Java 277, 379
Javascript 85–86, 187, 358, 380
Jockers, Matthew 14, 137, 365
JotNot 306, 380
jQuery 196, 207, 342, 380
Karr, Larry 148, 380
Kinomatics Project 21, 166, 168–70, 172–75, 380
Lantern 16–17, 108, 112–16, 120, 123, 216–17, 249, 359–60, 380

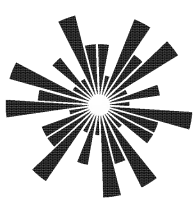
LDA *see* Dirichlet allocation
Levine, Elana H. 22–23
Library of Congress 36, 48, 77, 108, 336, 338, 349, 375
Long, Derek 7, 21, 24, 352, 365
low-tech digital 21, 24, 133, 141
MALLET 225, 227–28, 380
map 32, 60, 66, 68
    heat 110–11, 113, 379
    interactive 75, 83, 86, 377, 382–83
    Story Maps 85–86, 386
    see also digital; Google
MapHook 72
mapping 20–21, 57, 71–72, 76, 79–86, 113, 208, 256
    deep 70
    history 68–69
    mind-mapping software 302, 314, 315
    see also digital
Mapping Movies 69, 381
Mapping the City in Film 71, 381
Margaret Herrick Library 353, 364
Markov Chain 74, 381
McPherson, Tara 9–10
Media Ecology Project (MEP) 7, 23, 150, 249, 335–37, 340–41
media historiography 8, 10, 20, 296, 316, 348
media history 2, 10–13, 16–17, 19–24, 34, 45, 48, 59, 103, 137, 147–48, 151, 159, 220, 296, 309, 336–37, 348, 351, 361, 363, 374
Media History Digital Library (MHDL) 16–19, 22, 39, 48, 58, 103–04 108–09, 116–17, 120–21, 123–125, 216–17, 224–25, 231, 238–39, 249, 349, 354, 360, 382
MediaCommons 348, 379, 381
Mediathread 337, 341, 381
MEP *see* Media Ecology Project
Metadata *see* data
Meyers, Cynthia B. 1, 23, 381
MHDL *see* Media History Digital Library

MIRC *see* Moving Image Research Collections
Moretti, Franco 67, 273, 365, 377
Moving Image Research Collections (MIRC) 338, 354, 382 *see also* University of South Carolina
MP4 303, 382
multi-sited cinema *see* cinema
MySQL 171, 358 382 *see also* SQL
national cinema *see* cinema
National Press Club 49
Neatline 79, 84, 382
NEH Office of Digital Humanities 2
New Cinema History 68, 166, 168, 178
Newscape 339
newspaper 20, 31, 35–36, 40, 47–59, 69, 85, 171, 351–52, 374–75, 378
N-gram 22, 218, 220–24
Ngram Viewer 16, 116, 220, 382 *see also* Google
nontheatrical 2, 17, 39, 46, 50, 54–55, 57, 125
  cinema *see* cinema
  circuit 103
  exhibition 49, 106
  film 20, 46, 48, 60, 134
  filmmaking 112, 114, 118, 140
  magazines 17, 114
  screening 47, 50, 52
  spaces 32, 106
Northeast Historic Film 339, 355
OCR *see* Optical Character Recognition
Omeka 84, 351, 358
Onomy 337, 341, 383
Open Archives Initiative 383
Optical Character Recognition (OCR) 16, 18, 48, 149, 219, 302, 306–07, 382
ORBIS: The Stanford Geospatial Network Model of the Roman World 76, 383
Orlando Project 82
Paper Print Collection 337

Peabody Awards Collection 339
PeopleMovin 77
Perl 150, 383
Phan, Michelle 22, 197–98, 200, 202–03, 208
Photoshop *see* Adobe
PHP 187, 358, 383
Pierce, David 17, 146, 349, 354
Posner, Miriam 9, 21, 24, 358
Prelinger, Rick 134, 136, 141, 378
Premiere ProCC *see* Adobe Premiere ProCC
Project Arclight *see* Arclight
Project Bamboo 340, 383
Python 86, 366, 384
QuickTime Player 7 275, 384
radio 13, 39, 103, 110, 311, 320, 344, 353, 356
Red Hen Lab 339, 384
research strategy 273, 298, 304
ripping 197, 200, 384
Ruby on Rails 359, 384
SAT *see* Semantic Annotation Tool
Scalar 9, 84, 187–88, 337, 341, 361, 384
Scaled Entity Search (SES) 1, 19–20, 103–04, 108, 112, 118, 125, 363–65, 385
Scapple 314–15, 384
scripts 23, 196, 305, 313, 366, 367, 385
Scrivener 310–314, 385
Semantic Annotation Tool (SAT) 341
SES *see* Scaled Entity Search
Shotlogger 7
simulacrum 272–73
Slave Revolt in Jamaica, 1760–1761: A Cartographic Narrative 79
slice 22, 184, 186, 190–91, 270–71, 275–81, 296
small data *see* data
soap opera 22, 196, 301–03
Society for Cinema and Media Studies 338, 345, 357, 379
Software Studies Initiative 277, 385

Solr 15, 358, 360, 385
spatial humanities 65, 67, 74, 80, 83
Spehr, Paul 146–50,153–54, 352
Spiro, Lisa 14, 22
SQL *see* Structured Query Language
stack 22, 201, 204, 277–78, 281, 305, 313, 367, 385
STAT *see* Text Simple Text Analysis Tool
stop words/list 220–22, 227–28, 386
string literals 152, 386
Structured Query Language (SQL) 152, 176, 385 *see also* MySQL
tagging 326, 341, 382, 386
taxonomies 175, 340, 383
Taylor, Fraser 66–68, 78
Texas Archive of the Moving Image 4, 356
text-mining 4, 196
TextDNA 113, 386
TextSTAT (Simple Text Analysis Tool) 208, 386
topic modeling 14, 45, 137, 217–18, 223–28, 231, 365, 380, 386
trade papers 17, 22, 112, 347, 364–65
Tran, Tony 7, 21–23, 365
Tumblr 1, 6, 23, 187, 318, 323–328, 331, 387
    @Bcast_Md 325
    Klappersacks 328
    National Archives 325
UCLA Film and Television Archive 335, 338–39
University of Baltimore 339
University of Georgia 339 *see also* Walter J. Brown Media Archives and Peabody Awards Collection
University of Kansas 59
University of Kentucky 47
University of Maine 34, 387
University of Maryland-College Park 325
University of Missouri 49
University of Redlands 74
University of Richmond 79

University of South Carolina 5, 338–39, 351, 354, 382
University of Virginia 67, 84–85, 339
University of Wisconsin–Madison 13, 249
UNIX 154, 158, 387
useful
    cinema 139
    media 12, 103, 125, 135
Utrecht University 86, 350
Vectors 9, 360-61
VEMI Lab *see* Virtual Environment and Multimodal Interaction Laboratory
Verhoeven, Deb 21, 24, 68, 73, 81
Video Analysis Tableau 387
video essay 8, 9, 348, 379
Virtual Environment and Multimodal Interaction Laboratory (VEMI Lab) 342, 387
VisualHub 387
Visualizing Emancipation 79
Voltaire 75–76
Voyant 227, 387
W3C Open Annotation 340, 342, 387
Waller, Gregory A. 20–21, 392
Walter J. Brown Media Archives 339
Wasson, Haidee 1, 14, 20, 117
WFPP *see* Women Film Pioneers Project
Wiki 71, 75, 388
Williams, Mark 7, 23, 197
Women Film Pioneers Project (WFPP) 356, 388
WordPress 187, 361, 388
xQuery 152, 154, 158, 388
YouTube 1–2, 7–8, 22, 36, 196–200, 203, 206–07, 380
Zatoichi 282–84, 286, 288, 290–91, 294–95
Zotero 224, 388
Z-Projections 204–06, 209, 211, 288, 296


ARCLIGHT